GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 21, 2003, 23:30:28 ; Search time 31532 Seconds
(without alignments)
11771.313 Million cell updates/sec

Title: US-09-831-000-1_COPY_60966_70038
Perfect score: 9073
Sequence: 1 atggagagttccgtcggatg.....gcggccacggatcgccttga 9073

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 20454813386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 300 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		9073	100.0	133719	14	AF083501	AF083501 Macaca mu
2		3823	42.1	130733	14	AF210726	AF210726 Macaca mu
3		1665	18.4	1945	14	AF241163	AF241163 Cercopith
4	c	25	0.3	187157	9	AL139384	AL139384 Human DNA
5	c	24	0.3	143959	2	AC090055	AC090055 Oryza sat
6		23	0.3	90079	2	AC144923	AC144923 Mus muscu
7		23	0.3	105183	10	AC124702	AC124702 Mus muscu
8		23	0.3	164778	9	AP003394	AP003394 Homo sapi
9		23	0.3	179068	10	AC131111	AC131111 Mus muscu
10		23	0.3	182975	9	AP003558	AP003558 Homo sapi
11		23	0.3	188596	2	AC027247	AC027247 Homo sapi
12		23	0.3	188922	2	AC084833	AC084833 Homo sapi
13		23	0.3	191109	9	AP004249	AP004249 Homo sapi
14		23	0.3	195779	2	AC128738	AC128738 Mus muscu
15	c	23	0.3	195779	2	AC128738	AC128738 Mus muscu
16		23	0.3	198788	8	ATCHRIV46	AL161546 Arabidops
17		23	0.3	207674	8	ATFCA8	Z97343 Arabidopsis
18		23	0.3	227486	2	AC129867	AC129867 Rattus no
19	c	23	0.3	278353	2	AC106122	AC106122 Rattus no
20	c	22	0.2	1879	10	MMU249416	AJ249416 Mus muscu
21	c	22	0.2	12590	6	AX346822	AX346822 Sequence
22		22	0.2	21590	8	AP006134	AP006134 Lotus jap
23	c	22	0.2	41951	2	AC139177	AC139177 Ciona sav
24	c	22	0.2	77322	2	AC021334	AC021334 Homo sapi
25	c	22	0.2	110000	2	AC108356_2	Continuation (3 of
26		22	0.2	146761	2	AC132894	AC132894 Mus muscu
27		22	0.2	151506	10	AC126427	AC126427 Mus muscu
28	c	22	0.2	155852	2	AC135841	AC135841 Canis fam
29	c	22	0.2	167615	2	AC136889	AC136889 Canis fam
30		22	0.2	167910	2	AC023203	AC023203 Homo sapi
31		22	0.2	169159	2	AC023035	AC023035 Homo sapi
32	c	22	0.2	169332	2	AC126437	AC126437 Mus muscu
33		22	0.2	174012	5	AL929503	AL929503 Zebrafish
34		22	0.2	178201	9	AC025888	AC025888 Homo sapi
35		22	0.2	180213	2	AC120424	AC120424 Mus muscu
36		22	0.2	187559	2	AC137678	AC137678 Mus muscu
37	c	22	0.2	187685	2	AC126269	AC126269 Mus muscu
38	c	22	0.2	191804	2	AC019017	AC019017 Homo sapi
39		22	0.2	193454	2	AC084310	AC084310 Mus muscu
40	c	22	0.2	194247	2	BX005411	BX005411 Danio rer
41	c	22	0.2	194620	2	AC125520	AC125520 Mus muscu
42		22	0.2	195041	10	AC125195	AC125195 Mus muscu
43	c	22	0.2	195997	2	AC140374	AC140374 Mus muscu
44		22	0.2	196742	2	AC102808	AC102808 Mus muscu
45		22	0.2	198298	2	AC140314	AC140314 Mus muscu
46	c	22	0.2	203073	10	AC122322	AC122322 Mus muscu
47		22	0.2	205379	2	AC132141	AC132141 Mus muscu
48		22	0.2	206309	10	AL773582	AL773582 Mouse DNA
49	c	22	0.2	212140	2	AC122918	AC122918 Mus muscu
50		22	0.2	225220	2	AC131319	AC131319 Mus muscu
51	c	22	0.2	225439	2	AC104177	AC104177 Rattus no
52		22	0.2	237418	2	AC118928	AC118928 Mus muscu
53	c	22	0.2	242309	2	AC098167	AC098167 Rattus no
54	c	22	0.2	244698	2	AC105871	AC105871 Rattus no
55		22	0.2	251448	3	AE014819	AE014819 Plasmodiu
56		22	0.2	252749	2	AC096611	AC096611 Rattus no
57		22	0.2	267093	2	AC109408	AC109408 Rattus no
58		21	0.2	845	11	BV028873	BV028873 S212P6333
59	c	21	0.2	3628	6	AX598771	AX598771 Sequence
60	c	21	0.2	3628	6	AX598917	AX598917 Sequence
61	c	21	0.2	6127	6	AX345495	AX345495 Sequence
62		21	0.2	10910	1	ECO518835	AJ518835 Escherich
63	c	21	0.2	23481	9	AC110783	AC110783 Homo sapi
64	c	21	0.2	33501	2	AC014548	AC014548 Drosophil
65		21	0.2	34718	14	AF206304	AF206304 Human her

c	66	21	0.2	54618	9	AC104800 AC104800 Homo sapi
	67	21	0.2	68882	2	AC101569 AC101569 Mus muscu
c	68	21	0.2	68882	2	AC101569 AC101569 Mus muscu
c	69	21	0.2	69113	9	AL359960 AL359960 Human DNA
	70	21	0.2	74508	5	AL591511 AL591511 Zebrafish
	71	21	0.2	83544	8	AB025615 AB025615 Arabidops
c	72	21	0.2	85948	10	AL831757 AL831757 Mouse DNA
c	73	21	0.2	106793	9	AC002538 AC002538 Homo sapi
	74	21	0.2	109349	2	AC096839_7 Continuation (8 of
	75	21	0.2	110000	2	AC096839_2 Continuation (3 of
c	76	21	0.2	110000	2	AC096862_2 Continuation (3 of
c	77	21	0.2	110000	2	AC118188_1 Continuation (2 of
	78	21	0.2	110000	2	AC122235_3 Continuation (4 of
c	79	21	0.2	110000	2	AC127191_2 Continuation (3 of
c	80	21	0.2	118998	2	AC130200 AC130200 Medicago
	81	21	0.2	124884	6	AR283580 AR283580 Sequence
	82	21	0.2	124884	6	BD097170 BD097170 Method fo
	83	21	0.2	124884	14	HEVZVXX X04370 Varicella-Z
	84	21	0.2	125078	14	AB097932 AB097932 Human her
	85	21	0.2	125125	14	AB097933 AB097933 Human her
	86	21	0.2	125157	6	BD097171 BD097171 Method fo
c	87	21	0.2	128091	2	AC139626 AC139626 Takifugu
	88	21	0.2	129016	2	AC130257 AC130257 Rattus no
	89	21	0.2	129039	10	AL808113 AL808113 Mouse DNA
	90	21	0.2	131366	9	AL354701 AL354701 Human DNA
	91	21	0.2	139407	2	AL929509 AL929509 Danio rer
c	92	21	0.2	140488	2	AP004124 AP004124 Oryza sat
c	93	21	0.2	149849	9	AL590678 AL590678 Human DNA
	94	21	0.2	150965	9	AC091915 AC091915 Homo sapi
c	95	21	0.2	151236	9	HS404K8 AL023883 Human DNA
	96	21	0.2	151428	8	AC104322 AC104322 Oryza sat
	97	21	0.2	154279	9	AC112229 AC112229 Homo sapi
	98	21	0.2	154606	9	AC097358 AC097358 Homo sapi
c	99	21	0.2	156083	9	AC074087 AC074087 Homo sapi
	100	21	0.2	156690	2	AC027163 AC027163 Homo sapi
c	101	21	0.2	157519	8	AP001389 AP001389 Oryza sat
	102	21	0.2	159997	2	AC137496 AC137496 Homo sapi
c	103	21	0.2	160191	2	AC124963 AC124963 Medicago
	104	21	0.2	161417	9	CNS01RHX AL162632 Human chr
	105	21	0.2	162547	2	AC022172 AC022172 Homo sapi
	106	21	0.2	162752	10	AC121961 AC121961 Mus muscu
c	107	21	0.2	163265	2	BX511036 BX511036 Danio rer
c	108	21	0.2	163537	2	AC128956 AC128956 Rattus no
c	109	21	0.2	164076	9	AC008006 AC008006 Homo sapi
	110	21	0.2	164383	2	AC068023 AC068023 Homo sapi
c	111	21	0.2	165933	1	AP004174 AP004174 Mycoplasm
	112	21	0.2	166418	2	AC020990 AC020990 Homo sapi
	113	21	0.2	168862	9	AC090532 AC090532 Homo sapi
c	114	21	0.2	169303	9	AL354808 AL354808 Human DNA
c	115	21	0.2	169540	9	AC009248 AC009248 Homo sapi
	116	21	0.2	170085	2	AC119187 AC119187 Mus muscu
c	117	21	0.2	171168	2	AP005385 AP005385 Oryza sat
	118	21	0.2	171708	9	AC067956 AC067956 Homo sapi
c	119	21	0.2	171934	9	AC092050 AC092050 Homo sapi
c	120	21	0.2	173154	2	AC019132 AC019132 Homo sapi
c	121	21	0.2	178127	2	AC079076 AC079076 Homo sapi
c	122	21	0.2	179877	2	AC102051 AC102051 Mus muscu
	123	21	0.2	180338	2	AC137957 AC137957 Mus muscu
	124	21	0.2	185132	9	CNS01DUB AL133238 Human chr
c	125	21	0.2	187793	3	AC013247 AC013247 Drosophil
c	126	21	0.2	192541	2	AC079165 AC079165 Mus muscu
c	127	21	0.2	193772	2	AC016833 AC016833 Homo sapi
c	128	21	0.2	198181	2	AC035248 AC035248 Mus muscu
c	129	21	0.2	201309	10	AL669857 AL669857 Mouse DNA
	130	21	0.2	203021	2	AC122228 AC122228 Mus muscu
	131	21	0.2	206770	2	AC107763 AC107763 Mus muscu
c	132	21	0.2	212819	2	AC026550 AC026550 Homo sapi
	133	21	0.2	215819	9	AC140490 AC140490 Homo sapi
	134	21	0.2	218846	2	AC137416 AC137416 Rattus no
	135	21	0.2	221847	2	AC127929 AC127929 Rattus no
	136	21	0.2	223729	2	AC120905 AC120905 Rattus no
	137	21	0.2	225663	10	AC126428 AC126428 Mus muscu
c	138	21	0.2	231524	2	AC115181 AC115181 Rattus no
	139	21	0.2	233835	2	AC132778 AC132778 Rattus no
	140	21	0.2	239810	2	AC111901 AC111901 Rattus no
c	141	21	0.2	241787	2	AC115348 AC115348 Rattus no
c	142	21	0.2	242570	2	BX322570 BX322570 Danio rer
c	143	21	0.2	250768	2	AC128442 AC128442 Rattus no
c	144	21	0.2	252843	2	AC096261 AC096261 Rattus no
	145	21	0.2	260625	2	AC096018 AC096018 Rattus no
	146	21	0.2	262694	2	AC096503 AC096503 Rattus no
	147	21	0.2	265928	2	AC093937 AC093937 Rattus no
c	148	21	0.2	275331	2	AC119323 AC119323 Rattus no
c	149	21	0.2	284300	2	AC111967 AC111967 Rattus no
	150	21	0.2	291648	2	AC111639 AC111639 Rattus no
	151	21	0.2	294609	3	AE003787 AE003787 Drosophil
	152	21	0.2	297724	2	AC125901 AC125901 Rattus no
	153	21	0.2	300092	2	AC106230 AC106230 Rattus no
	154	21	0.2	312677	8	AE017068 AE017068 Oryza sat
c	155	21	0.2	322950	2	AC119791 AC119791 Rattus no
c	156	21	0.2	347950	1	AP003013 AP003013 Mesorhizo
c	157	20	0.2	255	6	AX439740 AX439740 Sequence
	158	20	0.2	356	6	AX438071 AX438071 Sequence
c	159	20	0.2	425	14	AF049081 AF049081 Simian fo
	160	20	0.2	461	6	AX408176 AX408176 Sequence
c	161	20	0.2	571	11	G98503 G98503 S208P6049FD
	162	20	0.2	581	10	MMU23096 U23096 Mus musculu
c	163	20	0.2	597	6	BD115583 BD115583 EST and e
c	164	20	0.2	618	8	AF388813 AF388813 Cortinari
c	165	20	0.2	636	8	AF114451 AF114451 Hygrophor
	166	20	0.2	714	8	AF488526 AF488526 Utricular
	167	20	0.2	736	6	BD019087 BD019087 Novel gen
	168	20	0.2	736	6	BD099025 BD099025 Novel gen
c	169	20	0.2	738	11	BV018811 BV018811 S212P6045
	170	20	0.2	774	6	AX721144 AX721144 Sequence
	171	20	0.2	797	8	AF482566 AF482566 Utricular
	172	20	0.2	798	8	AF482584 AF482584 Utricular
	173	20	0.2	819	8	AF482589 AF482589 Utricular
	174	20	0.2	829	8	AF482596 AF482596 Utricular
	175	20	0.2	832	8	AF482579 AF482579 Utricular
	176	20	0.2	834	6	AX349563 AX349563 Sequence
	177	20	0.2	834	8	AF482576 AF482576 Utricular
	178	20	0.2	834	8	AF482587 AF482587 Utricular
	179	20	0.2	837	8	AF482590 AF482590 Utricular
	180	20	0.2	843	8	AF482599 AF482599 Utricular
	181	20	0.2	844	8	AF482562 AF482562 Utricular
	182	20	0.2	846	8	AF482561 AF482561 Utricular
c	183	20	0.2	963	6	AX439053 AX439053 Sequence
	184	20	0.2	1019	9	HSM803193 AL831871 Homo sapi
c	185	20	0.2	1077	9	HSA325717 AJ325717 Homo sapi
c	186	20	0.2	1472	6	BD170613 BD170613 NF-kappa
c	187	20	0.2	1772	14	AY131198 AY131198 Human her
c	188	20	0.2	1901	9	AK094386 AK094386 Homo sapi
c	189	20	0.2	1913	9	BC015012 BC015012 Homo sapi
c	190	20	0.2	1924	6	BD170614 BD170614 NF-kappa
c	191	20	0.2	1925	6	AX059851 AX059851 Sequence
c	192	20	0.2	1927	6	AX136093 AX136093 Sequence
c	193	20	0.2	1927	6	BD123499 BD123499 Secretory
c	194	20	0.2	1927	9	AK075340 AK075340 Homo sapi
c	195	20	0.2	1933	9	AB028926 AB028926 Homo sapi
c	196	20	0.2	1977	6	AX376436 AX376436 Sequence
c	197	20	0.2	1977	6	AX454740 AX454740 Sequence
c	198	20	0.2	1977	6	AX491218 AX491218 Sequence
c	199	20	0.2	1977	6	AX574518 AX574518 Sequence
	200	20	0.2	2766	9	BC039903 BC039903 Homo sapi
c	201	20	0.2	2938	3	AB041522 AB041522 Lymnaea s
c	202	20	0.2	4261	1	AF538722 AF538722 Streptoco
c	203	20	0.2	8339	9	AL162912 AL162912 Human DNA
c	204	20	0.2	9238	6	AX323751 AX323751 Sequence
	205	20	0.2	10029	1	AE007526 AE007526 Clostridi
c	206	20	0.2	11131	6	AX281163 AX281163 Sequence
c	207	20	0.2	11131	6	AX345655 AX345655 Sequence
c	208	20	0.2	11131	6	AX356402 AX356402 Sequence
c	209	20	0.2	11327	6	AX251185 AX251185 Sequence
	210	20	0.2	15356	1	AE004867 AE004867 Pseudomon
	211	20	0.2	16932	1	AE001661 AE001661 Chlamydia

c 212	20	0.2	20215	1	AE002264	AE002264 Chlamydop
c 213	20	0.2	23793	3	U88178	U88178 Caenorhabdi
214	20	0.2	26994	8	AP006083	AP006083 Lotus jap
215	20	0.2	32392	6	AX392737	AX392737 Sequence
216	20	0.2	35100	6	AR065850	AR065850 Sequence
217	20	0.2	35100	6	AR127848	AR127848 Sequence
218	20	0.2	35100	6	AR194750	AR194750 Sequence
c 219	20	0.2	35638	9	AC112771	AC112771 Homo sapi
c 220	20	0.2	36361	3	LMFL4326	AL121861 Leishmani
c 221	20	0.2	41397	3	CEB0285	Z34533 Caenorhabdi
222	20	0.2	48081	2	AC014896	AC014896 Drosophil
c 223	20	0.2	53676	2	AC090184	AC090184 Homo sapi
c 224	20	0.2	56385	2	AC100596	AC100596 Mus muscu
225	20	0.2	57448	2	AC013539	AC013539 Homo sapi
c 226	20	0.2	60101	2	AC101124	AC101124 Mus muscu
c 227	20	0.2	61562	2	AC102798	AC102798 Homo sapi
228	20	0.2	62868	2	AC016799	AC016799 Homo sapi
229	20	0.2	63152	2	AC129559	AC129559 Mus muscu
230	20	0.2	65229	9	AL672108	AL672108 Human DNA
c 231	20	0.2	68591	2	AC106851	AC106851 Homo sapi
c 232	20	0.2	69201	9	AC004109	AC004109 Homo sapi
233	20	0.2	70116	2	AC016414	AC016414 Homo sapi
234	20	0.2	71145	2	AC016088	AC016088 Homo sapi
c 235	20	0.2	71271	2	AC102800	AC102800 Homo sapi
c 236	20	0.2	72576	10	BX005243	BX005243 Mouse DNA
237	20	0.2	86692	9	AC007527	AC007527 Homo sapi
238	20	0.2	87571	2	AC021900	AC021900 Homo sapi
c 239	20	0.2	87894	2	AC125031	AC125031 Mus muscu
240	20	0.2	90450	10	AL627226	AL627226 Mouse DNA
c 241	20	0.2	90452	2	AC100310	AC100310 Mus muscu
242	20	0.2	90826	9	AC008433	AC008433 Homo sapi
243	20	0.2	91495	9	AL596323	AL596323 Human DNA
c 244	20	0.2	94362	8	AC005405	AC005405 BAC F18A1
c 245	20	0.2	100000	9	AP000080	AP000080 Homo sapi
246	20	0.2	100464	9	AC106705	AC106705 Homo sapi
c 247	20	0.2	101996	2	AC135463	AC135463 Medicago
248	20	0.2	103737	9	AL603966	AL603966 Human DNA
c 249	20	0.2	105087	9	AL512343	AL512343 Human DNA
c 250	20	0.2	107458	9	AC023492	AC023492 Homo sapi
c 251	20	0.2	108044	9	AL360220	AL360220 Human DNA
252	20	0.2	108409	14	AF083424	AF083424 Ateline h
c 253	20	0.2	108895	9	AL137783	AL137783 Human DNA
254	20	0.2	109637	9	AC084032	AC084032 Homo sapi
c 255	20	0.2	110000	2	AC121331_1	Continuation (2 of
c 256	20	0.2	110000	2	AC126130_0	AC126130 Rattus no
257	20	0.2	110000	2	AC138524_0	AC138524 Homo sapi
258	20	0.2	110000	2	RN52A06_01	Continuation (2 of
259	20	0.2	110000	2	RN52A06_02	Continuation (3 of
260	20	0.2	110000	6	AR310754_08	Continuation (9 of
261	20	0.2	110000	6	AR310754_09	Continuation (10 o
262	20	0.2	110283	9	AC093245	AC093245 Homo sapi
263	20	0.2	112533	9	AL139097	AL139097 Human DNA
264	20	0.2	113027	14	SHE410493	AJ410493 Saimiriin
c 265	20	0.2	113778	2	AC109438	AC109438 Homo sapi
266	20	0.2	114377	2	AC143993	AC143993 Macaca mu
267	20	0.2	114540	9	AC107426	AC107426 Homo sapi
268	20	0.2	118724	9	AC105074	AC105074 Homo sapi
269	20	0.2	121014	9	AC022723	AC022723 Homo sapi
270	20	0.2	121287	9	AC010332	AC010332 Homo sapi
c 271	20	0.2	121336	9	AC006458	AC006458 Homo sapi
c 272	20	0.2	122132	9	AC102799	AC102799 Homo sapi
273	20	0.2	123018	2	AC102488	AC102488 Mus muscu
c 274	20	0.2	123521	9	HS75G13	AJ006998 Homo sapi
275	20	0.2	123962	2	AC133583	AC133583 Mus muscu
c 276	20	0.2	124102	9	AC021148	AC021148 Homo sapi
277	20	0.2	125041	2	AC068693	AC068693 Homo sapi
c 278	20	0.2	128757	2	AC142568	AC142568 Felis cat
c 279	20	0.2	129203	9	AC104819	AC104819 Homo sapi
c 280	20	0.2	129272	9	AL451125	AL451125 Human DNA
281	20	0.2	129601	2	AP005808	AP005808 Oryza sat
c 282	20	0.2	130427	10	AC012302	AC012302 Mus muscu
283	20	0.2	130891	2	AL354954	AL354954 Homo sapi
c 284	20	0.2	130922	2	AC101831	AC101831 Mus muscu
c 285	20	0.2	131140	2	AC027474	AC027474 Homo sapi
286	20	0.2	131811	9	AC068305	AC068305 Homo sapi
c 287	20	0.2	133309	9	AL355003	AL355003 Human DNA
288	20	0.2	133661	14	U93872	U93872 Kaposi's sa
c 289	20	0.2	133877	2	AC106922	AC106922 Rattus no
290	20	0.2	135345	9	AL356294	AL356294 Human DNA
c 291	20	0.2	135734	2	BX321868	BX321868 Danio rer
292	20	0.2	136448	2	AC012231	AC012231 Homo sapi
293	20	0.2	137508	14	KSU75698	U75698 Kaposi's sa
c 294	20	0.2	137806	2	AC115832	AC115832 Mus muscu
c 295	20	0.2	138003	2	AC018346	AC018346 Homo sapi
c 296	20	0.2	139526	2	AC143320	AC143320 Macaca mu
297	20	0.2	139635	9	AC013290	AC013290 Homo sapi
c 298	20	0.2	140120	9	AL390237	AL390237 Human DNA
c 299	20	0.2	140409	2	AC044825	AC044825 Homo sapi
300	20	0.2	142105	9	AC012374	AC012374 Homo sapi

ALIGNMENTS

RESULT 1
AF083501
LOCUS AF083501 133719 bp DNA linear VRL 26-JUN-2000
DEFINITION Macaca mulatta rhadinovirus 17577, complete genome.
ACCESSION AF083501
VERSION AF083501.3 GI:8714565
KEYWORDS .
SOURCE Macaca mulatta rhadinovirus 17577
ORGANISM Macaca mulatta rhadinovirus 17577
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Rhadinovirus.
REFERENCE 1 (bases 1 to 133719)
AUTHORS Searles,R.P., Bergquam,E.P., Axthelm,M.K. and Wong,S.W.
TITLE Sequence and genomic analysis of a Rhesus macaque rhadinovirus with similarity to Kaposi's sarcoma-associated herpesvirus/human herpesvirus 8
JOURNAL J. Virol. 73 (4), 3040-3053 (1999)
MEDLINE 99174001
PUBMED 10074154
REFERENCE 2 (bases 1 to 133719)
AUTHORS Searles,R.P., Bergquam,E.P., Axthelm,M.K. and Wong,S.W.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-1998) Division of Pathobiology and Immunology, Oregon Health Sciences University/ Oregon Regional Primate Research Center, 505 NW I85th Avenue, Beaverton, OR 97006, USA
REFERENCE 3 (bases 1 to 133719)
AUTHORS Searles,R.P., Bergquam,E.P., Axthelm,M.K. and Wong,S.W.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-1999) Division of Pathobiology and Immunology, Oregon Health Sciences University/ Oregon Regional Primate Research Center, 505 NW I85th Avenue, Beaverton, OR 97006, USA
REMARK Sequence update by submitter
REFERENCE 4 (bases 1 to 133719)
AUTHORS Searles,R.P., Bergquam,E.P., Axthelm,M.K. and Wong,S.W.
TITLE Direct Submission
JOURNAL Submitted (26-JUN-2000) Division of Pathobiology and Immunology, Oregon Health Sciences University/ Oregon Regional Primate Research Center, 505 NW I85th Avenue, Beaverton, OR 97006, USA
REMARK Sequence update by submitter
COMMENT On Jun 26, 2000 this sequence version replaced gi:5508840.
FEATURES Location/Qualifiers
source 1. .133719
/organism="Macaca mulatta rhadinovirus 17577"
/mol_type="genomic DNA"
/db_xref="taxon:83534"
repeat_region 1. .840
/note="partial terminal repeat"
CDS 1353. .2624
/codon_start=1
/product="R1"
/protein_id="AAD21330.1"
/db_xref="GI:4494908"

/translation="MFVLVLFMLLQPVSVELLPAKLTSVPTWCPPHPGDTYLLTCRGT
STARDQRSTQWFRNNTLMRGSNFYGRLVSVTPNATISDRYACQTKTTTRSNNIDFRVS
SSRLTLQERCSSYGYTYANNTRVLRCYSGGNVTLRNVVFHLNGTAVINGTTTNIHTFV
LTEKTGGTYFCSAFIGNEKFYSQTINVFFTSFTFKPTNDIPNESHFNKTGQIQQTASV
QHPENYVVFSVPVFSIGVLTGIAISLIMCWLFTIRCNENSESSTNSYASQTSYIQPSH
NQRSNTNECSRHTYRNAHQEESIEELPNQHTSETDSCCQLVLLEVKNVAYDGPQENTI
NEVMEQYDDVVVKNIEQTSYEDNVEHMDYSDTINPNFNYYSGLILEEVDEVFYNELEN
QYHGLILENLDHNEYNHLNELNMIEQYDWLE"

CDS complement(2692. .3258)
/note="ORF 2; similar to Kaposi's sarcoma-associated herpesvirus ORF 2"
/codon_start=1
/product="dihydrofolate reductase"
/protein_id="AAD21331.1"
/db_xref="GI:4494909"
/translation="MDITVNCIVAVDEQLGIGKNGTMPWPYLRNEMMYFQKMTSTPSV
VGEKNVVIMGKRTWFSIPEKKRPLVNRINIILSRELREPPHGAHFLARTLDDAFNFYR
QYKLKEQLNTVWVIGGKSVYESVLNYKCPLKLYITRIMESFDCDVFFPSINFTEYTML
SEIPGKDTNFEENGIKYKFQVYEKNFNK"

CDS 3676. .5613
/note="ORF 4; similar to Kaposi's sarcoma-associated herpesvirus ORF 4"
/codon_start=1
/product="complement binding protein"
/protein_id="AAD21332.1"
/db_xref="GI:4494910"
/translation="MTFKLFPLFLLHAIMYVHCDENCKPPHFTEYRVKSNTEKDLYSV
GETAELICRPGYVTNTKIITTECLQNGTWSTPNFPCDRKRCPTPADLLNGAVHIHGGD
NALKFGSNISYECNEGYDLIGSNVRFCILQDTENVNWDSNEPVCEIQKCIKPPAVEHG
DYLPNQDVYNYGDAITFKCSLSYTLVGSTTLVCTSNKKWSNSFPTCLMLVCESPQIDN
GYIDIGLSRRYNHGQSITVKCSDGYNIVGPETLTCTNTTWVPPLPKCVLVTNNPSTPM
PETPMPETPTPDYQKINLSTAKTATTPNAFVTTVVSPEKDDVTCVKPHFERFMVKAEN
DKEKYSVGASVELICRPGFTKMQSTVSVECLSNGTWTAPNAKCHRKKCPTPQELLNGE
YIVTSGEDAFKYGTNITYKCNEGYQLLGSMVRICMLKDDLKTVDWEPKAPICDIEKCK
PPPQITNGKYHPVKDFYQYLDTVTFSCNRDFSLVGDEMTTCISNTWNKPFPRCEQITC
SAPNIAHGKLLTGSSSVYKYGQSVTIGCETGFTLIGSEISTCKDSSWDPPLPTCVPAV
SMPSDTPKPETKKPNTPTPEAPKPNTPNVGTHTPFKPPPQNPPIAPPMSKWKRHVVLV
LFASVASLLFVLAALYCCFLK"

CDS 6045. .9443
/note="ORF 6; similar to Kaposi's sarcoma-associated herpesvirus ORF 6"
/codon_start=1
/product="ssDNA binding protein"
/protein_id="AAD21333.1"
/db_xref="GI:4494911"
/translation="MASKGNAGQPLEDNQGSRAPIGACGYVYAYSKQDFPFAEASILG
NRPSGSGVFSLPILYGLTVEHEFPLTVKAAYKKVDTTTLAVKVTCFHREVIVFHNASL
FRPVFDGTGLNELCEEARALFGYTQFIEPGPPHSIWNPLECPQLPDKDEMFLGVVVTE
GFKERLWRGCLVPAVFQTQQVQIAGRQAFKVPLYDEDLFAPHGHRMPRFYHKDVSAYL
YDSLFTSIAQALRLKDVTAVIHATEKQFMQDHYKIAKIVQAKQFSTTLPKTTDGSSHM
IVDSVVAELALSYGCMFLECPQDACELLNYDSWPIFDGCDSPEARVNALERWSAEQAV
HVAGQLFAANSVLYLTKVQKQAPRGQKGDVNVYNSFFLQHGLGFLNEATIKENGSEAF
KGVPSNALDGSSFTPYHLAYAASFSPHLLAKLCYYMQFLQHHKSSTNQAFNMVHYVGT
AANSEMCTLCHGNTPATCLNTLFYRLKDRFPAVTTPQRRDPYVVTGTAGTFNDLEILG
NFASFRDREEDGNPADEHPKYTYWQLCQTVTEKLSAIGITEDHDNHVNLITNIQSFLR
VFKGIDSIVDGEVMKFVNSMIKNNFNFREHVKSVHHILQFCCNVYWQAPCAVFLNLYY
KSLLWIIQDICLPYCMIYEQDNPAMGILPSEWLKMHFQTLWTNFKAACLDRGVLTGCE
LKIVHRDMFCDFFDTDAGSNGLMAPFKMQVRIARAMMVVPKSIKIKNRIIFSNTAGSE
AVQSGFVKPTGTRDTYVVAGPYMKFLNSLHRALFPDTKTAALYLWHKISQTNKTPVLK
DVPDDELAELVSYVKTNSLAFEETNVLDVVPDSLMSYARIKLNGAILRACGQIQFYAT
TLHCLTPVLQTIDAEEYPHVLGSAAIATPVAYLAEIRGRTALTVQTTARQPVAATGRL
RPVITVPMVVNKYTGVNGNNNVFHCGNLGYFAGRGVDRNLWPESSPFKKTGVSAMLRK
RHVMMTPIIDRLIKRAAGQTISTFEAESVKRSVQALLEDKDNPNLLKSVILELIRHLG
KGCQDLSSEDVQYYLGDYCMLTDEVLFTLDNIAQSGVPWTIEDAGALIEDRQDADDLQ
FVDSDDIATASCQPPEEQLPTPSAGALLAGKKRKINALLSDLDL"

CDS 9468. .11528
/note="ORF 7; similar to Kaposi's sarcoma-associated herpesvirus ORF 7"
/codon_start=1
/product="transport protein"
/protein_id="AAD21334.1"
/db_xref="GI:4494912"
/translation="MARELAALYAQLSALAVDLSLVIFADPRSIDGARILKTKTQIEN
LNRDLLPLLREQNSVETSSLSLEVEHLAKNIEDKLGELERSLRQRYSSREHFETLHLR
PECHYHSTVTFQFYGGGLIDVNMCLINDVELLCKRLGSVFYCIGANEALSGLNRVLTF
LSTLRGISPIPHPDLYVTSVPCVQCLREIELVPNQGSSLLAVLADRHCDHLCKKVRAE
PIHGLFETELSQLGLKVTKRSDATQHGVRSSADQLRESSLAAIQDHNIFKRVSASIME
LSNLIYWNAGQTGLQTGTENECSQMARLLTHEADMHEHRALITPKLSATHFYDCFRPD
PIESLFCGGLFNSIDDTINALSRDCSVTFFQQANYTNVMRKQNELFTRLNSILRQGSA
GSQKPATPSEPRTTTVAATAASDVIKDAQYRKEQYMKKVARDGFKKLTECLQTQSAVL
ANALCMRVWGGVAYGEASELVNHFLLRRRFVALPWEARCRSDQILFENSKYIKNSLYS
QRLSREHVEIITLQFYGLITGPLTRQSDLFPGPANVALAQCFEAAGMLPHHKMLVSEM
IWPQIQPKDWIDQTFNRFYQLPEGDLNAVQKSAWCFIRELVLSVALYNRTWEKTLRIF
SLAREKLSISNLDVKGLTSGLYLTYEQDAPLVLISQNTGWIFKDLYALLYHHLQLSDG
HDDN"

CDS 11515. .14004
/note="ORF 8; similar to Kaposi's sarcoma-associated herpesvirus ORF 8"
/codon_start=1
/product="glycoprotein B"
/protein_id="AAD21335.1"
/db_xref="GI:4494913"
/translation="MMITNRTRRLLRAWVVIIAIGTAVGENVTTPKGATTTAKPTPGP
STPTPPENPPRAEAFKFRVCSASATGELFRFNLEKTCPGTEDKTHQEGILMVFKKNIV
PHIFKVRRYRKVATSVTVYRGWTETAVTGKQEVIRPVPQYEINHMDTTYQCFSSMRVN
VNGIVNTYTDRDFTNQTVFLQPVEGLTDNIQRYFSQPVLYTTPGWFPGIYRVRTTVNC
EIVDMIARSAEPYSYFVTALGDTVEVSPFCHNDSTCSVAEKTENGLGARVLTNYTIVD
FATRQPTTETRVFADSGEYTVSWKAEDPKSAVCALTLWKTFPRAIQTTHEASYHFVAN
DVTATFTSPLSQVTNFTGTYPCLNDVIQKTLNATIKKLSDTHATNGSEQYYETEGGLF
LLWQPLTPLSLADEMRELNGTTPAPPTTTSTANRVRRSVGTNEQATDDLAAPQLQFAY
DKLRASINKVLEELSRAWCREQVRDTYMWYELSKINPTSVMTAIYGRPVSAKFVGDAI
SVTDCVAVDQASVSIHKSLRTSTPGICYSRPPVTFRFLNSTTLFKGQLGPRNEIILTD
NQVEACKETCEHYFIASNVTYYYKDYVFVKKINTSEISTLGTFIALNLSFIENIDFRV
IELYSRAEKKLSGSVFDIETMFREYNYYTQRLAGLREDLDNTIDLNRDRLARDLSEIV
ADLGDVGRTVVNVASSVITLFGSIVSGFINFIKSPFGGMLMILVIVAVVLIVFALNRR
TNAIAQAPIRMIYPDIDKMQPSGGKVDQEQIKNILAGMHQLQQEERRRLDEQQRSAPS
LFRRASDGLKRRFRGYKPLENEEAQEYEMSK"

CDS 14122. .17166
/note="ORF 9; similar to Kaposi's sarcoma-associated herpesvirus ORF 9"
/codon_start=1
/product="DNA polymerase"
/protein_id="AAD21336.1"
/db_xref="GI:4494914"
/translation="MDFFNPYLGPRGPRPHSHRGTDAPAPAGAGAVQPPPDVCRLIPA
CLRTPGAGGMIPVTIPFPPTYFENGARGDVLLANERSMWTARDRKPVAPDPQDQSITF
HAYDVVETTYAADRCAEVPSRFQTDIIPSGTVLKLLGRTEDGTSVCVNVFRQQVYFYA
KVPAGINVTHILQQALKNTAGRAACGFSTRRVNKRILKTYDVAEHPVTEITLSSGSML
STLSDRLVACGCEVFESNVDAVRRFVLDHGFTTFGWYSCARATPRLAARDARTALEFD
CSWEDLSVQADRSDWPPYRIVAFDIECTGEAGFPCATRDGDAVIQISCVFYTTREGAP
NPPNILFSVGTCDPIPDTDVLEFPSEYDMLVSFFAMIRDFEVDFLTGYNISNFDLPYL
ITRASQVYNLRLNEYTKIKTGSIFEVHEPRGGGGGFMRSVSKIKIAGIVPIDMYQVCR
EKLSLSDYKLDTVARQCLGGKKEDVSYKDIPPLFRSGPGGRAKVGSYCVMDSVLVMDL
LKMFMIHVEISEIAKLAKIQARRVLTDGQQLRVFSCLLEAAARENFILPVPTPEGQGG
YQGATVINPIPGFYDEPVLVVDFASLYPSIIQAHNLCYSTMIHGRDLHLHPNLTPDDY
ETFVLSGGPVHFVKKHKRESLLGRLLTVWLEKRRAIRRTLAACDDPSLKTILDKQQLA
IKVTCNAVYGFTGVASGLLPCINIAETVTLRGRTMLEMSKSYVEALTTEDLRTRLGRE
VTARHGARFRVVYGDTDSLFIACDGYSAEAVSAFCDDLAARITADLFPPPIKLEAEKT
FKCLLLLTKKRYIGVLLNDKMVMKGVDLIRKTACKFVQERCRAILDLVLHDPEVKAAA
RLLCKRPPHAVYEEGLPAGFIKIVEVLNASYLDLRNSVVPIEQLTFSTELSRPVCDYK
TTNLPHLAVYQKLASRCEELPQVHDRIPYVFVDAPGSLKSDLAEHPDYVRQHQIPVAV
DLYFDKLVHGAANILQCLFGNNADTTVAILYNFLNVPYKLFS"

CDS 17261. .18511
/note="ORF 10; similar to Kaposi's sarcoma-associated herpesvirus ORF 10"
/codon_start=1
/product="unknown"
/protein_id="AAD21337.1"
/db_xref="GI:4494915"
/translation="MLVNELSVVLGDWEVTFHRGRFSFVNLTRLQTFKGHGGYARVRL
PFSLDQLLHQHFAFGLVTRLKELPPFSDCVALIAPLDSGGDADAARVAPGFVLDSSRP
LTVWVNASGRHTIRFCLLFLKPIDLERAVTYVFGENGGARSEGTPKPTCATESLPGGP
LRVSGEASQTSPHSFVAYFPTANSVACLSLLRLQVRPFSDDAAHRDARISPKYVTFSN
SGGNVCKASVHTLSPSRCKTAQMEIIYAPGDPNAEIVLGQSGPVLPTHTGGRVLGVYA
DAEKTIQPGSSAEVRVQLIFQQGAAARGDLAFLVTGVAPEPLFVVTPALLLSGCTTHL
RLFNPNGTPTTIKRDTLVAAAAPCPVVRLSSADDAPRDLVASPDTGALSINAFTIPVG

FPGVVSAECHVSLRDNGVHERMNH"
CDS 18520. .19749
/note="ORF 11; similar to Kaposi's sarcoma-associated herpesvirus ORF 11"

Query Match 100.0%; Score 9073; DB 14; Length 133719;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 9073; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGGAGAGTTCCGTCGGATGGACCAAACACGTCGAACCAAATCCGGGGTTCATCTTGAAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60966 ATGGAGAGTTCCGTCGGATGGACCAAACACGTCGAACCAAATCCGGGGTTCATCTTGAAC 61025

Qy         61 ATGACGTCCGATGCCAAAGTCAGGGGTGTCGTGGATCACGTCAGTCGCCTGTCAAATATA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61026 ATGACGTCCGATGCCAAAGTCAGGGGTGTCGTGGATCACGTCAGTCGCCTGTCAAATATA 61085

Qy        121 ACTACCAGCCCACCGGAAATGGGTTGGTACGACCTGGCCTTCGATCCGGCTGAAGACTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61086 ACTACCAGCCCACCGGAAATGGGTTGGTACGACCTGGCCTTCGATCCGGCTGAAGACTCC 61145

Qy        181 GGGCCGTTCTTGCCGTTTACCGTTTATCTAATTACGGGAACTGCTGGTGCTGGGAAAAGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61146 GGGCCGTTCTTGCCGTTTACCGTTTATCTAATTACGGGAACTGCTGGTGCTGGGAAAAGT 61205

Qy        241 ACCAGCATATCGGCCCTGTACCAAAATTTAAACTGCCTGATCACGGGCGCGACCACCATA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61206 ACCAGCATATCGGCCCTGTACCAAAATTTAAACTGCCTGATCACGGGCGCGACCACCATA 61265

Qy        301 GCCGCACAGAACCTATCGCGTCGCCTAAAGACGTTCTGTCCCACGATCTTCAGCGCTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61266 GCCGCACAGAACCTATCGCGTCGCCTAAAGACGTTCTGTCCCACGATCTTCAGCGCTTTT 61325

Qy        361 GGCTTTAAGAGCCGACACATCAATATAGCCGTCAGAAAAGCTCATCAGACCGGAGCCGTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61326 GGCTTTAAGAGCCGACACATCAATATAGCCGTCAGAAAAGCTCATCAGACCGGAGCCGTA 61385

Qy        421 TCCATAGAGCAAATTCAGCAACAGGAGCTATCGAAGTATTGGCCGGTTATAGTGGACATT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61386 TCCATAGAGCAAATTCAGCAACAGGAGCTATCGAAGTATTGGCCGGTTATAGTGGACATT 61445

Qy        481 ATGAAAGAGGTTATGGCGAAAAAACCCAATGGCATGTACGGGACTATATCCAACGCGAAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61446 ATGAAAGAGGTTATGGCGAAAAAACCCAATGGCATGTACGGGACTATATCCAACGCGAAT 61505

Qy        541 TTTGAAACCCTCTCGAGAATGACCGGACCGTGTTTATGGACTTCCAATATTATTGTAATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61506 TTTGAAACCCTCTCGAGAATGACCGGACCGTGTTTATGGACTTCCAATATTATTGTAATC 61565

Qy        601 GACGAGGCCGGAACCCTGTCCTCTTACATACTCACCACCGTCGTGTTCTTTTACTGGTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61566 GACGAGGCCGGAACCCTGTCCTCTTACATACTCACCACCGTCGTGTTCTTTTACTGGTTC 61625

Qy        661 CTAAACAGCTGGCTAAATACCCCTCTTTACCGCCAGGGGGCGGTTCCGTGCATAGTATGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61626 CTAAACAGCTGGCTAAATACCCCTCTTTACCGCCAGGGGGCGGTTCCGTGCATAGTATGC 61685

Qy        721 GTCGGTTCGCCAACGCAGACAAACGCGTTCCAGTCAACTTACAACCACGGGACGCAAAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61686 GTCGGTTCGCCAACGCAGACAAACGCGTTCCAGTCAACTTACAACCACGGGACGCAAAAG 61745

Qy        781 ACGGAGATATCGTCGTGCGAGAACATCCTAACATTCATGATAGGGAAGAAGGTCGTATCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61746 ACGGAGATATCGTCGTGCGAGAACATCCTAACATTCATGATAGGGAAGAAGGTCGTATCT 61805

Qy        841 GAGTACGTACACTTGGAGAGAAACTGGGCGCTGTTTATAAACAACAAGCGCTGCACCGAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61806 GAGTACGTACACTTGGAGAGAAACTGGGCGCTGTTTATAAACAACAAGCGCTGCACCGAT 61865

Qy        901 CTGCAGTTCGGACACCTTCTAAAAATTTTAGAGTATAATCTTCCCATTCCTGACGAAGTC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61866 CTGCAGTTCGGACACCTTCTAAAAATTTTAGAGTATAATCTTCCCATTCCTGACGAAGTC 61925

Qy        961 ATGAGTTACGTAGACAGATTTGTCGTTCCTAAAAGTAAGATTATGGATCCTTTAGAATAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61926 ATGAGTTACGTAGACAGATTTGTCGTTCCTAAAAGTAAGATTATGGATCCTTTAGAATAC 61985

Qy       1021 ATTGGCTGGACCCGACTCTTTTTGTCACATAGCGAGGTAAAGGCGTATCTAACAAACCTA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61986 ATTGGCTGGACCCGACTCTTTTTGTCACATAGCGAGGTAAAGGCGTATCTAACAAACCTA 62045

Qy       1081 CACACATGTCTAACGCTAGGGGGCGATACCAGGGACACGAAGCTCTTTACCTGTCCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62046 CACACATGTCTAACGCTAGGGGGCGATACCAGGGACACGAAGCTCTTTACCTGTCCCGTG 62105

Qy       1141 GTGTGCGAGGTGTTTGTGAAGCCGTTTGAGGAATACAAGCGGGCCGTCAACCTCACCAAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62106 GTGTGCGAGGTGTTTGTGAAGCCGTTTGAGGAATACAAGCGGGCCGTCAACCTCACCAAC 62165

Qy       1201 CTCACCGTGACCGAGTGGGTGACAAAAAACCTCTTTAAGTTAAGTAATTATTCGCAGTTT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62166 CTCACCGTGACCGAGTGGGTGACAAAAAACCTCTTTAAGTTAAGTAATTATTCGCAGTTT 62225

Qy       1261 GTGGACCAGGACATGTCCATAGTCGCCACGGAATCCACTGAACGTTCGACCCAGGTTACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62226 GTGGACCAGGACATGTCCATAGTCGCCACGGAATCCACTGAACGTTCGACCCAGGTTACC 62285

Qy       1321 TTTATCACCAAGTTTGTTAAAAACAGCCACGTGTCTTTAAATGGAAAAACAAAAAAATGC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62286 TTTATCACCAAGTTTGTTAAAAACAGCCACGTGTCTTTAAATGGAAAAACAAAAAAATGC 62345

Qy       1381 ATATGTGGGTTTCAGGGTACGTACTTTGAGTTCAAAAGAATCCTAGACAGCGAACTCTTC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62346 ATATGTGGGTTTCAGGGTACGTACTTTGAGTTCAAAAGAATCCTAGACAGCGAACTCTTC 62405

Qy       1441 GTGGAGACGCATTCGCAAGACCGTCCGGAATACGTATACGGTTTCTTAAACACGCTACTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62406 GTGGAGACGCATTCGCAAGACCGTCCGGAATACGTATACGGTTTCTTAAACACGCTACTG 62465

Qy       1501 TACAACGCCATGTACTCGTTTCACGCGTACGGCGTGACTAGGTCACACGAGAAATATCTG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62466 TACAACGCCATGTACTCGTTTCACGCGTACGGCGTGACTAGGTCACACGAGAAATATCTG 62525

Qy       1561 CAAGACCTAAAATTTGCACCCCTCCCGGCCGCTCTGGCAACCGGGCGTGTAGACCTTCAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62526 CAAGACCTAAAATTTGCACCCCTCCCGGCCGCTCTGGCAACCGGGCGTGTAGACCTTCAA 62585

Qy       1621 ACGGTTCGTGAAGAGTTAAACCTGGAAGACGACATCTTCTACCACGTGTGCAGTCCCCCG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62586 ACGGTTCGTGAAGAGTTAAACCTGGAAGACGACATCTTCTACCACGTGTGCAGTCCCCCG 62645

Qy       1681 CCACCCGCGGGTATCACCTCCCTCCAGGTTTTGGTCGACACGTACTGCGCCCTAAAGGAC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62646 CCACCCGCGGGTATCACCTCCCTCCAGGTTTTGGTCGACACGTACTGCGCCCTAAAGGAC 62705

Qy       1741 GTGTTCGCCTCCAGAATAAAGGTGGCGTGTCGCTGGTTTGGCGGGGAGTTTGAGAAGGAA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62706 GTGTTCGCCTCCAGAATAAAGGTGGCGTGTCGCTGGTTTGGCGGGGAGTTTGAGAAGGAA 62765

Qy       1801 ACGTTTTCCGCGTTTACGGTTAACATGGTCGTAAGGGACGGAGTTGACTTTGTCTCCCCT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62766 ACGTTTTCCGCGTTTACGGTTAACATGGTCGTAAGGGACGGAGTTGACTTTGTCTCCCCT 62825

Qy       1861 TCAGAACGTCTCAACGGGCTGTTGGCGTTTGCATCGACCGTTGAATCGTATAAAATTAAG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62826 TCAGAACGTCTCAACGGGCTGTTGGCGTTTGCATCGACCGTTGAATCGTATAAAATTAAG 62885

Qy       1921 GGGTACACGTTTTTACCGGTAGCGTTCGGTCGCTGTCAGGGTTTGCCACTCAGTGATGAC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62886 GGGTACACGTTTTTACCGGTAGCGTTCGGTCGCTGTCAGGGTTTGCCACTCAGTGATGAC 62945

Qy       1981 CTCAGGAAGAAGATGCCCTCCCTGGTCGTGCAGGACTCTAGCGGTTTTATCGCGTGCCTA 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62946 CTCAGGAAGAAGATGCCCTCCCTGGTCGTGCAGGACTCTAGCGGTTTTATCGCGTGCCTA 63005

Qy       2041 GAGAATAACATAACCAAATTGACCGAAACCATGGAGGACGGGAGCGTTTTCCAAGTGTGC 2100
```

```
Db      63006 GAGAATAACATAACCAAATTGACCGAAACCATGGAGGACGGGAGCGTTTTCCAAGTGTGC 63065
Qy       2101 TGTGCGGGGGACTATGGGGTCAGCTCAAATTTAGCCATGACCATCGTAAAGGCACAGGGA 2160
Db      63066 TGTGCGGGGGACTATGGGGTCAGCTCAAATTTAGCCATGACCATCGTAAAGGCACAGGGA 63125
Qy       2161 ATGTCGTTGGAGCGAGTAGCCGTAGTATTTGGGTCCCACAAGAACGTCCAGACAAGCCAC 2220
Db      63126 ATGTCGTTGGAGCGAGTAGCCGTAGTATTTGGGTCCCACAAGAACGTCCAGACAAGCCAC 63185
Qy       2221 GTGTATGTAGCAATATCAAGGGCTGTTAACTCAAATTATTTGGTCATGGACAGCAACCCC 2280
Db      63186 GTGTATGTAGCAATATCAAGGGCTGTTAACTCAAATTATTTGGTCATGGACAGCAACCCC 63245
Qy       2281 CTTAAAACCCTCCTCAGAGAACCAGTCGATAACACCTCCGCCAAGCATATAGTCCGCGCC 2340
Db      63246 CTTAAAACCCTCCTCAGAGAACCAGTCGATAACACCTCCGCCAAGCATATAGTCCGCGCC 63305
Qy       2341 CTCCACAACCCAAACACAACCCTCATCTACTAAATAAATAAAAACAAGCAAATGGAAAAC 2400
Db      63306 CTCCACAACCCAAACACAACCCTCATCTACTAAATAAATAAAAACAAGCAAATGGAAAAC 63365
Qy       2401 ATTGTGTTTTTATTCAGTCCAACCACGGCCACGGATAGTTGTCATTTCCACACACCGGGG 2460
Db      63366 ATTGTGTTTTTATTCAGTCCAACCACGGCCACGGATAGTTGTCATTTCCACACACCGGGG 63425
Qy       2461 GGGTGGCCGCCATAGTTTGACGACCAGAAACGCCCGGGCCCTCTGGGTTAGAAGTAGCCC 2520
Db      63426 GGGTGGCCGCCATAGTTTGACGACCAGAAACGCCCGGGCCCTCTGGGTTAGAAGTAGCCC 63485
Qy       2521 CGCCCCCCGTGGTCGAGGTCATACGTCTCTTAGGGCCCTGGGTTGTACCAACATAAATCA 2580
Db      63486 CGCCCCCCGTGGTCGAGGTCATACGTCTCTTAGGGCCCTGGGTTGTACCAACATAAATCA 63545
Qy       2581 CGTCGTCATCACCGTTGCAGTCCCTGCGTGACGCGCGCCTCTTGCAGGCGGCACGAACAC 2640
Db      63546 CGTCGTCATCACCGTTGCAGTCCCTGCGTGACGCGCGCCTCTTGCAGGCGGCACGAACAC 63605
Qy       2641 CGCCGCTTGTAACGTGATCTGCATCGTCGTCGCTGTCTGAAGATGAACACAGGTCAATCA 2700
Db      63606 CGCCGCTTGTAACGTGATCTGCATCGTCGTCGCTGTCTGAAGATGAACACAGGTCAATCA 63665
Qy       2701 CATCAGTGGCGCCCGTGCCCTCATCGGCGTCTGATTCGTCCCATGTCTCAACAGAGCCTC 2760
Db      63666 CATCAGTGGCGCCCGTGCCCTCATCGGCGTCTGATTCGTCCCATGTCTCAACAGAGCCTC 63725
Qy       2761 GGTCGTCCACTTCGTCGTCATCTGAAATAATAATGGTCTCTTGTCCTGCAGTTGATCCAG 2820
Db      63726 GGTCGTCCACTTCGTCGTCATCTGAAATAATAATGGTCTCTTGTCCTGCAGTTGATCCAG 63785
Qy       2821 CAACAACAGGTTCTTCTTCATCGGAAGACCATCCGCCCTCTGCCGGTGATCTAGATATAT 2880
Db      63786 CAACAACAGGTTCTTCTTCATCGGAAGACCATCCGCCCTCTGCCGGTGATCTAGATATAT 63845
Qy       2881 CACCTGACCAGGCGGGCGAGGCCCGTGGCGAAAAGCCGTCAGCTTCTCCGCCACTTTCAT 2940
Db      63846 CACCTGACCAGGCGGGCGAGGCCCGTGGCGAAAAGCCGTCAGCTTCTCCGCCACTTTCAT 63905
Qy       2941 CAGTGTCCTCGTAGTCATCTATAAAAAAATCGTCTTCGCTGTCTGATTCAGATGACTGGG 3000
Db      63906 CAGTGTCCTCGTAGTCATCTATAAAAAAATCGTCTTCGCTGTCTGATTCAGATGACTGGG 63965
Qy       3001 GTGCGCGGTGTTGGCCAGTAACCACCGCAACGGGCCGCGGACCACCAAACGCGGTCAGTG 3060
Db      63966 GTGCGCGGTGTTGGCCAGTAACCACCGCAACGGGCCGCGGACCACCAAACGCGGTCAGTG 64025
Qy       3061 GCGCGGGCGCGCTCGGTGGTGTGGTGGCACCTGTCGAAGTAATGACGGTAGGTAGCCGGC 3120
Db      64026 GCGCGGGCGCGCTCGGTGGTGTGGTGGCACCTGTCGAAGTAATGACGGTAGGTAGCCGGC 64085
Qy       3121 CGCTCTGGTTAAAGGACTCACCGTCCCGTGCGTGCGGATCTATTCTAGGTGGGGAGGAGA 3180
Db      64086 CGCTCTGGTTAAAGGACTCACCGTCCCGTGCGTGCGGATCTATTCTAGGTGGGGAGGAGA 64145
Qy       3181 CCGGCGGTTCGTCTTCTTCATCCACATCACTGTAGTCGCTGTTCTCCGAATGAGAGTCGG 3240
Db      64146 CCGGCGGTTCGTCTTCTTCATCCACATCACTGTAGTCGCTGTTCTCCGAATGAGAGTCGG 64205
Qy       3241 TTTCGTCTGGCGTTTTTGGAGTAGTGGGAGGGGAATCTTGGTCAAAAACATCCTCAGGTG 3300
Db      64206 TTTCGTCTGGCGTTTTTGGAGTAGTGGGAGGGGAATCTTGGTCAAAAACATCCTCAGGTG 64265
Qy       3301 CAGTCATGTCAATCACTGGGGGGCCTCCGTACCCGGGCGGTACCGGCAACCCCGTACGGG 3360
Db      64266 CAGTCATGTCAATCACTGGGGGGCCTCCGTACCCGGGCGGTACCGGCAACCCCGTACGGG 64325
Qy       3361 ACCCGGCGAACAAAAAACGTCTTGTTCTCTTTTTCCTAGGTGCCCCGGGAATCGGCAGCA 3420
Db      64326 ACCCGGCGAACAAAAAACGTCTTGTTCTCTTTTTCCTAGGTGCCCCGGGAATCGGCAGCA 64385
Qy       3421 TCCTGGGAGTGGCCGGCGGGGTCCTTGACGGGTCGGATAAGAACATAGCCATGGCCGAAA 3480
Db      64386 TCCTGGGAGTGGCCGGCGGGGTCCTTGACGGGTCGGATAAGAACATAGCCATGGCCGAAA 64445
Qy       3481 CGTTCACCTGTAAAACGCACTGCCTATCCCCGATCAGTTAATATTCCAGTCAACGGCACC 3540
Db      64446 CGTTCACCTGTAAAACGCACTGCCTATCCCCGATCAGTTAATATTCCAGTCAACGGCACC 64505
Qy       3541 CCTCCGATTCTGAACTAGATAGTCATTCGCAAGTTTAAAATGGTTGCAGCCTAGAAACTG 3600
Db      64506 CCTCCGATTCTGAACTAGATAGTCATTCGCAAGTTTAAAATGGTTGCAGCCTAGAAACTG 64565
Qy       3601 CGGCCAGGGAGAACCGGTTGCCGCGTGGGCGGTAGCCAGGGGAGACGGATGGCGCGCCTT 3660
Db      64566 CGGCCAGGGAGAACCGGTTGCCGCGTGGGCGGTAGCCAGGGGAGACGGATGGCGCGCCTT 64625
Qy       3661 GAGGACGAGATGTCTCTGTCCGTTTATCAGCGTCGCCTTGTCTATAGCCTTTCTTCCCCA 3720
Db      64626 GAGGACGAGATGTCTCTGTCCGTTTATCAGCGTCGCCTTGTCTATAGCCTTTCTTCCCCA 64685
Qy       3721 CAGCATAAAAACGCACCTCTGAAGCTTGGCAGAAAGGCAGCTTATTATGTAGCTGGTGAA 3780
Db      64686 CAGCATAAAAACGCACCTCTGAAGCTTGGCAGAAAGGCAGCTTATTATGTAGCTGGTGAA 64745
Qy       3781 CCAATCCCAGCCAAGGTTGGCGTGTGACCCCGCCTTCCCTCTTTCCACCGTCAGAATGGT 3840
Db      64746 CCAATCCCAGCCAAGGTTGGCGTGTGACCCCGCCTTCCCTCTTTCCACCGTCAGAATGGT 64805
Qy       3841 GTTTAAAAGCAGAACTCCCCGTTTTGCCCAACAGTCTAAGCACCCGTGAGAAGGAGCGGT 3900
Db      64806 GTTTAAAAGCAGAACTCCCCGTTTTGCCCAACAGTCTAAGCACCCGTGAGAAGGAGCGGT 64865
Qy       3901 GAACCCAGGTACAGTATTGGCTATCTCTTTAAAAATATTTTTGAGGCTTGGAGGTATACT 3960
Db      64866 GAACCCAGGTACAGTATTGGCTATCTCTTTAAAAATATTTTTGAGGCTTGGAGGTATACT 64925
Qy       3961 GTAGTCCGGAGCAACGCTAAAGGCTAGTCCGGTGGCTTGACCGCGATGGTAAGGATCCTG 4020
Db      64926 GTAGTCCGGAGCAACGCTAAAGGCTAGTCCGGTGGCTTGACCGCGATGGTAAGGATCCTG 64985
Qy       4021 GCCAAGGATCACGACTTTAATATCCTCCGGTTCGCAACAATACGACCACCACATAATTCT 4080
Db      64986 GCCAAGGATCACGACTTTAATATCCTCCGGTTCGCAACAATACGACCACCACATAATTCT 65045
Qy       4081 ATCTATCGGTGGATAAATCACGGTTACGTTGCTCATATCCATAACGCGCTTCAGCAGAGC 4140
Db      65046 ATCTATCGGTGGATAAATCACGGTTACGTTGCTCATATCCATAACGCGCTTCAGCAGAGC 65105
Qy       4141 CGCGAGCTTCTGTTTTAAAAATGGAGAAAGATTTAAAAATTCAAGCCATGAGTCGCTAAG 4200
Db      65106 CGCGAGCTTCTGTTTTAAAAATGGAGAAAGATTTAAAAATTCAAGCCATGAGTCGCTAAG 65165
Qy       4201 TAGCAACGTTTGTGTGGAGGGTTCTTCTAACACTTCTGGTGACATTTTACTCCAAACTAT 4260
Db      65166 TAGCAACGTTTGTGTGGAGGGTTCTTCTAACACTTCTGGTGACATTTTACTCCAAACTAT 65225
```

```
Qy       4261 TGTTTTAAGCCAACCCTCCATAAATCCCCGCGTCTGGTAATTACTTTACGTATACGTGAA 4320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65226 TGTTTTAAGCCAACCCTCCATAAATCCCCGCGTCTGGTAATTACTTTACGTATACGTGAA 65285

Qy       4321 ACTTTTGTACTTCTGTCTGAGAAATCCAACAGAGGTGGCCGCGTATAAGCAGTTCCGCTT 4380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65286 ACTTTTGTACTTCTGTCTGAGAAATCCAACAGAGGTGGCCGCGTATAAGCAGTTCCGCTT 65345

Qy       4381 TCACTACCGTTTATGTGAAAGTCTTCAAACTCGGCTAAAAACTCATCAAGACTATATTGC 4440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65346 TCACTACCGTTTATGTGAAAGTCTTCAAACTCGGCTAAAAACTCATCAAGACTATATTGC 65405

Qy       4441 ATATAATTCAACAATTGAAGCTCTTCTCTGGGGACACTGTAATTTAACTTGTACAAAATC 4500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65406 ATATAATTCAACAATTGAAGCTCTTCTCTGGGGACACTGTAATTTAACTTGTACAAAATC 65465

Qy       4501 CTTATAAAAAACGCCCTAAGGTGAAATCCATTTACACATATTTCTGTGGTAGAGCGATCA 4560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65466 CTTATAAAAAACGCCCTAAGGTGAAATCCATTTACACATATTTCTGTGGTAGAGCGATCA 65525

Qy       4561 GCTTTATATCGCAATGTTGCTGTACACAGCCCTGGACATAAAGCTAAATTTGTTAAATGA 4620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65526 GCTTTATATCGCAATGTTGCTGTACACAGCCCTGGACATAAAGCTAAATTTGTTAAATGA 65585

Qy       4621 AAAGAAGCAATCTCAAACGGCTCCGGAAATATAAAGTCACTAAGTGGTATTATTTTACAA 4680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65586 AAAGAAGCAATCTCAAACGGCTCCGGAAATATAAAGTCACTAAGTGGTATTATTTTACAA 65645

Qy       4681 CATGGATTTACTACAAGCCCGGTACAACATGTAATTAAAAAAAATCCACATGTAATAAAT 4740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65646 CATGGATTTACTACAAGCCCGGTACAACATGTAATTAAAAAAAATCCACATGTAATAAAT 65705

Qy       4741 AAACTTAAGGTATACATACTTCTCATGTTTTATACATAATATGAACGCTAATGCTGTACT 4800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65706 AAACTTAAGGTATACATACTTCTCATGTTTTATACATAATATGAACGCTAATGCTGTACT 65765

Qy       4801 TATATAATATGTATAGAATTATGCCAATACAGTCATGGTAAGTATATTTTAAGGCATAAT 4860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65766 TATATAATATGTATAGAATTATGCCAATACAGTCATGGTAAGTATATTTTAAGGCATAAT 65825

Qy       4861 GCAAACATCAAATAATATAACACAAAATGCACGCTTCCGGATATCCGCCCAGCTGTTAAC 4920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65826 GCAAACATCAAATAATATAACACAAAATGCACGCTTCCGGATATCCGCCCAGCTGTTAAC 65885

Qy       4921 CCGAAAATACAGAAATGACTACACAAACACACCTGAAACCAATTTTATTCTCAACATATG 4980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65886 CCGAAAATACAGAAATGACTACACAAACACACCTGAAACCAATTTTATTCTCAACATATG 65945

Qy       4981 CAACAATATTTCAGGGTAACACCATGTTAATAAAATACGCAGCATGCACATTTTTAGCTA 5040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65946 CAACAATATTTCAGGGTAACACCATGTTAATAAAATACGCAGCATGCACATTTTTAGCTA 66005

Qy       5041 AGCTCCTCAAAGACAATTTCCTCTTCAATTGATGACTCATCAACCTCGGTTTCCGTGTTA 5100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66006 AGCTCCTCAAAGACAATTTCCTCTTCAATTGATGACTCATCAACCTCGGTTTCCGTGTTA 66065

Qy       5101 ATGTCGGAAATATATGAGTCCAGGATACCCTCGTCATCGCCACAAATTAATTCCAGGGCT 5160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66066 ATGTCGGAAATATATGAGTCCAGGATACCCTCGTCATCGCCACAAATTAATTCCAGGGCT 66125

Qy       5161 TGTAAAATATCATCCAGTGAATCAGCTGCTAATGTCAAGGATGTTGTTTCTAATTTTAGC 5220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66126 TGTAAAATATCATCCAGTGAATCAGCTGCTAATGTCAAGGATGTTGTTTCTAATTTTAGC 66185

Qy       5221 CCATCCAAATGGTGAGTGGGTGACACTTTGGGACAAGACCCCGTAAAATTGGCACGCTCC 5280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66186 CCATCCAAATGGTGAGTGGGTGACACTTTGGGACAAGACCCCGTAAAATTGGCACGCTCC 66245

Qy       5281 ACACAAACGGAAGGATGCATTTGATCGTTCACGAGGGGCAGCATATTTATTTTTATAACA 5340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66246 ACACAAACGGAAGGATGCATTTGATCGTTCACGAGGGGCAGCATATTTATTTTTATAACA 66305

Qy       5341 TCAGGAACCGGACCCGTCAATGACGCACATATGTCCATTAAGATTGGAGTCGTGACACTG 5400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66306 TCAGGAACCGGACCCGTCAATGACGCACATATGTCCATTAAGATTGGAGTCGTGACACTG 66365

Qy       5401 GTTGAATTGAGGGTGAAAAAAAAATGTATTCGTGTCACATGCGTAACTATTTTTTAACCAC 5460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66366 GTTGAATTGACGGTGAAAAAAAATGTATTCGTGTCACATGCGTAACTATTTTTTAACCAC 66425

Qy       5461 ACGAGGGCGGAATAAAACGGGTCCCCGGCGCTAAGGCCGTGAATGCGGCAGGCACGGCTC 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66426 ACGAGGGCGGAATAAAACGGGTCCCCGGCGCTAAGGCCGTGAATGCGGCAGGCACGGCTC 66485

Qy       5521 ATTGCCATTTCCAGATTTGTTAAATCTGTTCGAACATACGAGTCTAAAATCAAGTATAAC 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66486 ATTGCCATTTCCAGATTTGTTAAATCTGTTCGAACATACGAGTCTAAAATCAAGTATAAC 66545

Qy       5581 CACGACAAGTTCAAACAGGAGGCAACGCGGCTGGTTTCCACCCGGACATCCCCGCTCACA 5640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66546 CACGACAAGTTCAAACAGGAGGCAACGCGGCTGGTTTCCACCCGGACATCCCCGCTCACA 66605

Qy       5641 AACCCGAACTCTCTAGGAACGCACATCAAAGCGTTCAGACAAAATTCCGAAACCGAAACC 5700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66606 AACCCGAACTCTCTAGGAACGCACATCAAAGCGTTCAGACAAAATTCCGAAACCGAAACC 66665

Qy       5701 CTAGTTCTGAGCACCATCTGGTTATAAACTGCGGTTAAGAGACGAGCGGCCACACACTGC 5760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66666 CTAGTTCTGAGCACCATCTGGTTATAAACTGCGGTTAAGAGACGAGCGGCCACACACTGC 66725

Qy       5761 TGCTTGTGAGGATTCAGCTTAGGTGGCCTGCAGGCTTGCTGACAGGCCCGTAGCTGCCTG 5820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66726 TGCTTGTGAGGATTCAGCTTAGGTGGCCTGCAGGCTTGCTGACAGGCCCGTAGCTGCCTG 66785

Qy       5821 GCGGCGCTCGCGCACCTTTGACAGCCGCACGCCAGTTCCAGAGCGAGGCAGAAGCGCTCT 5880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66786 GCGGCGCTCGCGCACCTTTGACAGCCGCACGCCAGTTCCAGAGCGAGGCAGAAGCGCTCT 66845

Qy       5881 TTGCAGTCGCGCCATATCTCGGCCATAGGAGATTCGCTCGCGTGATCTTTTAGGTGCATA 5940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66846 TTGCAGTCGCGCCATATCTCGGCCATAGGAGATTCGCTCGCGTGATCTTTTAGGTGCATA 66905

Qy       5941 TATATCTGTTGGGTAACAAACCTCACTCCCTGTAAAAGGGGAATAAGGTCCGTTCTTTCT 6000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66906 TATATCTGTTGGGTAACAAACCTCACTCCCTGTAAAAGGGGAATAAGGTCCGTTCTTTCT 66965

Qy       6001 ATCGTTTTCTCTGAATCTATAGTTACATCTTTAAGTATAACCAAAGACGCTAAAAACCCA 6060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66966 ATCGTTTTCTCTGAATCTATAGTTACATCTTTAAGTATAACCAAAGACGCTAAAAACCCA 67025

Qy       6061 GGATGGTCGACACCTTTAAAAAATCGCAAAAGCGACCGGATGGCTCTATCCGCGTTTCCA 6120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67026 GGATGGTCGACACCTTTAAAAAATCGCAAAAGCGACCGGATGGCTCTATCCGCGTTTCCA 67085

Qy       6121 TGTCGTTCAAAGGTAGTCACGATGGATCTCCAGTTAGACTCTGTCTCCCTGTTAACACCC 6180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67086 TGTCGTTCAAAGGTAGTCACGATGGATCTCCAGTTAGACTCTGTCTCCCTGTTAACACCC 67145

Qy       6181 TTAACTGGAATGGAAACAGCCATGACACCGTGAACTTCCTGATGTCTCCTAAAACTAACC 6240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67146 TTAACTGGAATGGAAACAGCCATGACACCGTGAACTTCCTGATGTCTCCTAAAACTAACC 67205

Qy       6241 CCGAAACAGAGCTAAATACCAATGACTGTCACCCCTACCCCAAGCCACGCCCCCGTACTA 6300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67206 CCGAAACAGAGCTAAATACCAATGACTGTCACCCCTACCCCAAGCCACGCCCCCGTACTA 67265

Qy       6301 TTAGACCAGGGTGAGTAACCACGCTATCCTTTAAAAACCCATACGTGGAGTTTGTAAGGT 6360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67266 TTAGACCAGGGTGAGTAACCACGCTATCCTTTAAAAACCCATACGTGGAGTTTGTAAGGT 67325

Qy       6361 AAACAGCTCGTGTATATTTCAGACGCCTGTCATAAAATGGATACCGACGACAATCAGGTA 6420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67326 AAACAGCTCGTGTATATTTCAGACGCCTGTCATAAAATGGATACCGACGACAATCAGGTA 67385

Qy       6421 ATTAAACTTTTTTTATTCAAGATTCAGGAGGGCGTGTTTACAGTGGTGTAGGTGGGAGCA 6480
```

```
Db      67386 ATTAAACTTTTTTTATTCAAGATTCAGGAGGGCGTGTTTACAGTGGTGTAGGTGGGAGCA 67445

Qy       6481 TATCTCGCCAATGGGAATGGCTGATGAATCCACACTTAGTGCTCGGATCAGGTGCTCTGT 6540

Db      67446 TATCTCGCCAATGGGAATGGCTGATGAATCCACACTTAGTGCTCGGATCAGGTGCTCTGT 67505

Qy       6541 GTACTTTACTAGTGCCGATGAACCTAAATTTGGCGTGTTTAGGTCGTACCAACCAAAAAA 6600

Db      67506 GTACTTTACTAGTGCCGATGAACCTAAATTTGGCGTGTTTAGGTCGTACCAACCAAAAAA 67565

Qy       6601 CTGCCTGGCCTCGCCTTTTAGAAGCTCTATGCAAGCTTTAATCGTGTCCAATAGCTTGTC 6660

Db      67566 CTGCCTGGCCTCGCCTTTTAGAAGCTCTATGCAAGCTTTAATCGTGTCCAATAGCTTGTC 67625

Qy       6661 TTTAAAAATGCAACTCTGGTACATCTTTATGACAGTGGTCCAAAAAAAACAAAGATTTAA 6720

Db      67626 TTTAAAAATGCAACTCTGGTACATCTTTATGACAGTGGTCCAAAAAAAACAAAGATTTAA 67685

Qy       6721 AAACACGTTAAACTGCGTGTCCTGGTAGTCCTCGTATATAACCTCCTCAACAAGAAAAAA 6780

Db      67686 AAACACGTTAAACTGCGTGTCCTGGTAGTCCTCGTATATAACCTCCTCAACAAGAAAAAA 67745

Qy       6781 TTTTTTAACCAAACTCGCCAGGTACTGAAACGATGGTACGGACAGGTCGTGAAAAGTGTC 6840

Db      67746 TTTTTTAACCAAACTCGCCAGGTACTGAAACGATGGTACGGACAGGTCGTGAAAAGTGTC 67805

Qy       6841 TATCATCTCCTTCACCTCCCCTTTGAAGGTTTTGGTTACACCGACCATGTGAGATAGGCA 6900

Db      67806 TATCATCTCCTTCACCTCCCCTTTGAAGGTTTTGGTTACACCGACCATGTGAGATAGGCA 67865

Qy       6901 CCAGCTCAGGGGGGAGGTCGGATCGTGAGGGGGTAACAATTCGTGGGCGTGGGGGTACTC 6960

Db      67866 CCAGCTCAGGGGGGAGGTCGGATCGTGAGGGGGTAACAATTCGTGGGCGTGGGGGTACTC 67925

Qy       6961 CAGTTTTAGGATCTCCGGCAGGACGCGTACGAGTTCTTCGTCAAATCCTACGCAGCCCGC 7020

Db      67926 CAGTTTTAGGATCTCCGGCAGGACGCGTACGAGTTCTTCGTCAAATCCTACGCAGCCCGC 67985

Qy       7021 ATCCACCAGGGGCAGTACTGAGTTAATAAGTCCCCGAATTCTATCGTTCGTAATTTTGTG 7080

Db      67986 ATCCACCAGGGGCAGTACTGAGTTAATAAGTCCCCGAATTCTATCGTTCGTAATTTTGTG 68045

Qy       7081 GAGCTCCTTTAGACAGTAAACCATGTTGCTCCGACACTGCGGTACGATGAACCGCTCTCC 7140

Db      68046 GAGCTCCTTTAGACAGTAAACCATGTTGCTCCGACACTGCGGTACGATGAACCGCTCTCC 68105

Qy       7141 TTGTCTGTGTTGCATGAGCACCTGGAGTATCTTAGACAGGTATAGGGCCCGTTCCACTCT 7200

Db      68106 TTGTCTGTGTTGCATGAGCACCTGGAGTATCTTAGACAGGTATAGGGCCCGTTCCACTCT 68165

Qy       7201 AAGGGTTGCCGCTTGGTAGAGCGGATTGCGGGCATTGACGGTCAGTGTGTTCATTTCACC 7260

Db      68166 AAGGGTTGCCGCTTGGTAGAGCGGATTGCGGGCATTGACGGTCAGTGTGTTCATTTCACC 68225

Qy       7261 GACTATCATGCCGAGTGGCGGATAGTGGAACTTGTACATATGATTTAAAAGATGGTCTTT 7320

Db      68226 GACTATCATGCCGAGTGGCGGATAGTGGAACTTGTACATATGATTTAAAAGATGGTCTTT 68285

Qy       7321 CCCGTAATGGCGACTCATGGTCGTGTTTTAACGTTCACAGGGTGTTAAAAAACTCAGAAG 7380

Db      68286 CCCGTAATGGCGACTCATGGTCGTGTTTTAACGTTCACAGGGTGTTAAAAAACTCAGAAG 68345

Qy       7381 GTCCTGCGTGGACAACTTTTGCGGACTCTCCGAACCGCTCAGGGATAAGCTGTACCACGC 7440

Db      68346 GTCCTGCGTGGACAACTTTTGCGGACTCTCCGAACCGCTCAGGGATAAGCTGTACCACGC 68405

Qy       7441 GGTGCTTCTTATAAACGAAGCATACGTCAAAAAACATGACCCACGTGACCTAAATAAGCT 7500

Db      68406 GGTGCTTCTTATAAACGAAGCATACGTCAAAAAACATGACCCACGTGACCTAAATAAGCT 68465

Qy       7501 CACAGAAGAAGTCTGTCGATGTATTGTAATGGAATGTGCCTCCCTGGGTCCCATCTCCGG 7560

Db      68466 CACAGAAGAAGTCTGTCGATGTATTGTAATGGAATGTGCCTCCCTGGGTCCCATCTCCGG 68525

Qy       7561 GCTCATCGCGGATCTGAATCTCTTTAACCTGTTCTGCCTCTACCGGGGTTCCCGGGTTAA 7620

Db      68526 GCTCATCGCGGATCTGAATCTCTTTAACCTGTTCTGCCTCTACCGGGGTTCCCGGGTTAA 68585

Qy       7621 AACCCGCGGGGCCGCCACCTGTAACGTCCCGTGCGCAGAATGCGCCCAGGGCATCGTGAG 7680

Db      68586 AACCCGCGGGGCCGCCACCTGTAACGTCCCGTGCGCAGAATGCGCCCAGGGCATCGTGAG 68645

Qy       7681 AATTCTCACCGAGAGGGCCCTTTGTTGTACCGAAAAGATGTTCATAGCGTCTGCCTGCAG 7740

Db      68646 AATTCTCACCGAGAGGGCCCTTTGTTGTACCGAAAAGATGTTCATAGCGTCTGCCTGCAG 68705

Qy       7741 CGGGGTCGTTATACCGCCCCAGTTGGCCAGAGTTCTGCACGACGTCTACGCGGAAATGAA 7800

Db      68706 CGGGGTCGTTATACCGCCCCAGTTGGCCAGAGTTCTGCACGACGTCTACGCGGAAATGAA 68765

Qy       7801 GGCCAAGTGTTTGGGGGCGTGGCGTCGTCTCATATGTTGCAGGCGGCCTATCATGGCTAT 7860

Db      68766 GGCCAAGTGTTTGGGGGCGTGGCGTCGTCTCATATGTTGCAGGCGGCCTATCATGGCTAT 68825

Qy       7861 TGCCGACTCGGTCCTCGTCACTTATAATACCCTGGATGCCGAGGGAAAACTGGAACTTAG 7920

Db      68826 TGCCGACTCGGTCCTCGTCACTTATAATACCCTGGATGCCGAGGGAAAACTGGAACTTAG 68885

Qy       7921 ACTCAAAGCACTGTGCAAACTTGTTTTTCAACCCATCTTTCTTCAAAGAATCTTAGCCCC 7980

Db      68886 ACTCAAAGCACTGTGCAAACTTGTTTTTCAACCCATCTTTCTTCAAAGAATCTTAGCCCC 68945

Qy       7981 TATGCAGCTACTGGCCAACGGGAAGATGGTTCCTGACAATTATTTTACCATCACCGGTAC 8040

Db      68946 TATGCAGCTACTGGCCAACGGGAAGATGGTTCCTGACAATTATTTTACCATCACCGGTAC 69005

Qy       8041 GGCCGAGAAGAGGCGCCCTGTCGTGACTGGAAGTACTAGCGGGATGACGTGTCCGGGAAG 8100

Db      69006 GGCCGAGAAGAGGCGCCCTGTCGTGACTGGAAGTACTAGCGGGATGACGTGTCCGGGAAG 69065

Qy       8101 CAGCCTTGTCCCCGATTCCTTAATCCTGCCAGTATGCGAACCGGGGGCTTCTCCCGGCACC 8160

Db      69066 CAGCCTTGTCCCCGATTCCTTAATCCTGCCAGTATGCGAACCGGGGGCTTCTCCCGGCACC 69125

Qy       8161 CCTGGTTGACCTCAGTAATGTCTTAGAAAATCCAGAAATCATCCTCAGCGCCCCACCCCT 8220

Db      69126 CCTGGTTGACCTCAGTAATGTCTTAGAAAATCCAGAAATCATCCTCAGCGCCCCACCCCT 69185

Qy       8221 GAGTCAATTTGTCATCACAAACACGCACCCCAGTCTGCCTCAGTCAGTCAGCATTATTAC 8280

Db      69186 GAGTCAATTTGTCATCACAAACACGCACCCCAGTCTGCCTCAGTCAGTCAGCATTATTAC 69245

Qy       8281 GCCAACCCAGGGCGTTGTTCCCGGCCAATGTTTTATGGACACGTGGAAAGCGGTGTCACA 8340

Db      69246 GCCAACCCAGGGCGTTGTTCCCGGCCAATGTTTTATGGACACGTGGAAAGCGGTGTCACA 69305

Qy       8341 GAGCATTCACCACCAGGCACAGACGCCTATTTTGGCCGCCGCACTAACCGGTTCGACATC 8400

Db      69306 GAGCATTCACCACCAGGCACAGACGCCTATTTTGGCCGCCGCACTAACCGGTTCGACATC 69365

Qy       8401 TGCGGCCCCTGGCCCGCATATCGCATGTTCCCCAGTTGCCGGCACGTCTCGGCAGGTGGA 8460

Db      69366 TGCGGCCCCTGGCCCGCATATCGCATGTTCCCCAGTTGCCGGCACGTCTCGGCAGGTGGA 69425

Qy       8461 AGGGTCCGCGGGCGTCGATTGCGGGAAACCAGCATGCGTTCCGCAGCCCGCGTTACCGCC 8520

Db      69426 AGGGTCCGCGGGCGTCGATTGCGGGAAACCAGCATGCGTTCCGCAGCCCGCGTTACCGCC 69485

Qy       8521 CAATGTCCCCGCCAAGAGGATGGAAACGGTAGCACAGTTGGGAAACGCTCCCGTAAAAAA 8580

Db      69486 CAATGTCCCCGCCAAGAGGATGGAAACGGTAGCACAGTTGGGAAACGCTCCCGTAAAAAA 69545

Qy       8581 CGTCCACATCGGAGGCCGCGTATACGCTCCACTGGTTAATATACCAATAATAGACTTAAC 8640

Db      69546 CGTCCACATCGGAGGCCGCGTATACGCTCCACTGGTTAATATACCAATAATAGACTTAAC 69605
```

```
Qy       8641 GTCCCCGTCAGGGTCCGGCCAGAGTCCGGCCGATATCGCCAACACTCCAGAGTCCCGCAT 8700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69606 GTCCCCGTCAGGGTCCGGCCAGAGTCCGGCCGATATCGCCAACACTCCAGAGTCCCGCAT 69665

Qy       8701 GGCGGCCGGCTCTCCGCCCTTCGCCGAAACCGCCGCAACGGTCCCCGCTAAGAGAAAGCA 8760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69666 GGCGGCCGGCTCTCCGCCCTTCGCCGAAACCGCCGCAACGGTCCCCGCTAAGAGAAAGCA 69725

Qy       8761 GCCACGCGAGGACGTGGCAGACAAAAGACTGAAGGGCGACGTTCGGGGCGCCGCAACAGT 8820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69726 GCCACGCGAGGACGTGGCAGACAAAAGACTGAAGGGCGACGTTCGGGGCGCCGCAACAGT 69785

Qy       8821 AAACCACCCTTTCCCGGGACCGTCCGGGATGCGCGTTCGCGAGCAGGGCTTATTCGATTT 8880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69786 AAACCACCCTTTCCCGGGACCGTCCGGGATGCGCGTTCGCGAGCAGGGCTTATTCGATTT 69845

Qy       8881 AATCGAAAGCTCCACGGATGTAACCGCGAACGCATCTGGACCCAAAAACGACGACGACAT 8940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69846 AATCGAAAGCTCCACGGATGTAACCGCGAACGCATCTGGACCCAAAAACGACGACGACAT 69905

Qy       8941 GCTAGCGGCTATCCTACAGGACCTGTATGGACTACAGTCCCCCCCGGCCATCGATTCCCC 9000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69906 GCTAGCGGCTATCCTACAGGACCTGTATGGACTACAGTCCCCCCCGGCCATCGATTCCCC 69965

Qy       9001 CTCCAGCAACTCGGACAATGAGGAGATATTTCCAGAGGTTAGTCCGCCATCTAGCGGCCA 9060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69966 CTCCAGCAACTCGGACAATGAGGAGATATTTCCAGAGGTTAGTCCGCCATCTAGCGGCCA 70025

Qy       9061 CGGATCGCCTTGA 9073
             |||||||||||||
Db      70026 CGGATCGCCTTGA 70038
```

```
RESULT 2
AF210726
LOCUS       AF210726               130733 bp    DNA     linear   VRL 27-MAR-2000
DEFINITION  Macaca mulatta rhadinovirus 26-95 long unique region L-DNA,
            complete sequence.
ACCESSION   AF210726
VERSION     AF210726.1  GI:7329990
KEYWORDS    .
SOURCE      Macaca mulatta rhadinovirus 26-95
  ORGANISM  Macaca mulatta rhadinovirus 26-95
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Rhadinovirus.
REFERENCE   1  (bases 1 to 130733)
  AUTHORS   Alexander,L., Denekamp,L., Knapp,A., Auerbach,M.R., Damania,B. and
            Desrosiers,R.C.
  TITLE     The primary sequence of rhesus monkey rhadinovirus isolate 26-95:
            sequence similarities to Kaposi's sarcoma-associated herpesvirus
            and rhesus monkey rhadinovirus isolate 17577
  JOURNAL   J. Virol. 74 (7), 3388-3398 (2000)
  MEDLINE   20173730
   PUBMED   10708456
REFERENCE   2  (bases 1 to 130733)
  AUTHORS   Alexander,L., Denekamp,L.M., Knapp,A., Auerbach,M., Czajak,S.,
            Damania,B. and Desrosiers,R.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-DEC-1999) Microbiology, New England Regional Primate
            Research Center, One Pinehill Dr, Southborough, MA 01772, USA
FEATURES             Location/Qualifiers
     source          1. .130733
                     /organism="Macaca mulatta rhadinovirus 26-95"
                     /mol_type="genomic DNA"
                     /isolate="Macaca mulatta rhadinovirus isolate 26-95"
                     /db_xref="taxon:119193"
     gene            513. .1784
                     /gene="R1"
     CDS             513. .1784
                     /gene="R1"
                     /codon_start=1
                     /product="ORFR1"
                     /protein_id="AAF59980.1"
                     /db_xref="GI:7329991"
                     /translation="MFVLVLFMLLQPVSVELLPAKLTSVPTWCPPHPGDTYLLTCRGT
                     STARDQRSTQWFRNNTLMRGSNFYGRLVSVTPNATISDRYACQTKTTTRSNNIDFRVS
                     SSRLTLQERCSSYGYTYANNTRVLRCYSGGNVTLRNVVFHLNGTAVINGTTRNIHTFV
                     LTEKTGGTYFCSAFIGNEKFYSQTINVFFTSFTFKPTNDIPNESHFNKTGQIQQTASV
                     QHPENYVVFSVPVFSIGVLTGIAISLIMCWLFTIRCNENSESSTNSYASQTSYIQPSH
                     NQRSNTNECSRHTYRNAHQEESIEELPNQHTSETDSCCQLVLLEVKNVAYDGPQENTI
                     NEVMEQYDDVVVENIEQTSYEDNVEHMDYSDTINPNFNYYSGLILEEVDEVFYNELEN
                     QYHGLILENLDHNEYNHLNELNMIEQYDWLE"
     gene            complement(1852. .2418)
                     /gene="ORF02"
     CDS             complement(1852. .2418)
                     /gene="ORF02"
                     /codon_start=1
                     /product="DHFR"
                     /protein_id="AAF59981.1"
                     /db_xref="GI:7329992"
                     /translation="MDIAVNCIVAVDEQLGIGKNGTMPWPYLRNEMMYFQKMTSTPSV
                     VGEKNVVIMGKRTWFSIPEKKRPLVNRINIILSRELREPPHGAHFLARTLDDAFNFYR
                     QYKLKEQLNTVWVIGGKSVYESVLNYKCPLKLYITRIMESFDCDVFFPSINFTEYTML
                     SELPGKDTNFEENGIKYKFQVYEKNFNK"
     gene            2595. .3782
                     /gene="ORF04"
     CDS             2595. .3782
                     /gene="ORF04"
                     /codon_start=1
                     /product="complement binding protein"
                     /protein_id="AAF59982.1"
                     /db_xref="GI:7329993"
                     /translation="MFWIPITLCFICQVILVDSKDGENVTCLKPPFDRFMVKTANQNE
                     NYAVGTRVELICRPGFYKLQANVYVECLSNGTWTTPNAECRRKRCSNPEDILNGEVII
                     TDSDNAFKFGSNITYKCNTGYLLLGATVRTCLLKYDSNLVDWQPAAPTCEIEKCKKQP
                     DIENGKYYPVQEFYNYLETITFTCNKDFSLIGNTTTTCMTNGTWSSPVPKCQQITCSA
                     PNIDHGTLLVGSSRVYKHGQSVTIGCEDGFTLNGHKMCTCEYSLWNPPLPTCVPINKT
                     DPTPSEVPSPGTKEQEMPTPENPKSHESETTTTETPKTQTHKSETPSPKVPNPETHKP
                     TTPEAGISKQTTTNRPSKAPSQNPPIAPPMSKWKRHVVLVLFASVASLLFVLAALYCC
                     FLK"
     gene            4213. .7611
                     /gene="ORF06"
     CDS             4213. .7611
                     /gene="ORF06"
                     /codon_start=1
                     /product="ssDBP"
                     /protein_id="AAF59983.1"
                     /db_xref="GI:7329994"
                     /translation="MASKGNAGQPLEDNQGSRAPIGACGYVYAYSKQDFPFAEASILG
                     NRPSGSGVFSLPILYGLTVEHEFPLTVKAAYKKVDTTTLAVKVTCFHREVIVFHNASL
                     FRPVFDGTGLNELCEEARALFGYTQFIEPGPPHSIWNPLECPQLPDKDEMFLGVVVTE
                     GFKERLWRGCLVPAVFQTQQVQIAGRQAFKVPLYDEDLFAPHGHRMPRFYHKDVSAYL
                     YDSLFTSIAQALRLKDVTAVIHATEKQFMQDHYKIAKIVQAKQFSTTLPKTTDGSSHM
                     IVDSVVAELALSYGCMFLECPQDACELLNYDSWPIFDGCDSPEARVNALERWSAEQAV
                     HVAGQLFAANSVLYLTKVQKQAPRGQKGDVNVYNSFFLQHGLGFLNEATIKENGSEAF
                     KGVPSNALDGSSFTPYHLAYAASFSPHLLAKLCYYMQFLQHHKSSTNQAFNMVHYVGT
                     AANSEMCTLCHGNTPATCLNTLFYRLKDRFPAVTTPQRRDPYVVTGTAGTFNDLEILG
                     NFASFRDREEDGNPADEHPKYTYWQLCQTVTEKLSAIGITEDHDNHVNLITNIQSFLR
                     VFKGIDSIVDGEVMKFVNSMIKNNFNFREHVKSVHHILQFCCNVYWQAPCAVFLNLYY
                     KSLLWIIQDICLPYCMIYEQDNPAMGILPSEWLKMHFQTLWTNFKAACLDRGVLTGCE
                     LKIVHRDMFCDFFDTDAGSNGLMAPFKMQVRIARAMMVVPKSIKIKNRIIFSNTAGSE
                     AVQSGFVKPTGTRDTYVVAGPYMKFLNSLHRALFPDTKTAALYLWHKISQTNKTPVLK
                     DVPDDELAELVSYVKTNSLAFEETNVLDVVPDSLMSYARIKLNGAILRACGQIQFYAT
                     TLHCLTPVLQTIDAEEYPHVLGSAAIATPVAYLAEIRGRTALTVQTTARQPVAATGRL
                     RPVITVPMVVNKYTGVNGNNNVFHCGNLGYFAGRGVDRNLWPESSPFKKTGVSAMLRK
                     RHVMMTPIIDRLIKRAAGQTISTFEAESVKRSVQALLEDKDNPNLLKSVILELIRHLG
                     KGCQDLSSEDVQYYLGDYCMLTDEVLFTLDNIAQSGVPWTIEDAGALIEDRQDADDLQ
                     FVDSDDIATASCQPPEEQLPTPSAGALLAGKKRKINALLSDLDL"
     gene            7636. .9696
                     /gene="ORF07"
     CDS             7636. .9696
                     /gene="ORF07"
                     /codon_start=1
                     /product="transport protein"
                     /protein_id="AAF59984.1"
```

```
                /db_xref="GI:7329995"
                /translation="MARELAALYAQLSALAVDLSLVIFADPRSIDGARILKTKTQIEN
                LNRDLLPLLREQNSVETSSLSLEVEHLAKNIEDKLGELERSLRQRYSSREHFETLHLR
                PECHYHSTVTFQFYGGGLIDVNMCLINDVELLCKRLGSVFYCIGANEALSGLDRVLTF
                LSTLRGISPIPHPDLYVTSVPCVQCLREIELVPNQGSSLLAVLADRHCDHLCKKVRAE
                PIHGLFETELSQLGLKVTKRSDATQHGVRSSADQLRESSLAAIQDHNIFKRVSASIME
                LSNLIYWNAGQTGLQTGTENECSQMARLLTHEADMHEHRALITPKLSATHFYDCFRPD
                PIESLFCGGLFNSIDDTINALSRDCSVTFFQQANYTNVMRKQNELFTRLNSILRQGSA
                GSQKPATPSEPRTTTVAATAASDVIKDAQYRKEQYMKKVARDGFKKLTECLQTQGAVL
                ANALCMRVWGGVAYGEASELVNHFLLRRRFVALPWEARCRSDQILFENSKYIKNSLYS
                QRLSREHVEIITLQFYGLITGPLTRQSDLFPGPANVVLAQCFEAAGMLPHHKMLVSEM
                IWPQIQPKDWIDQTFNRFYQLPEGDLNAVQKSAWCFIRELVLSVALYNRTWEKTLRIF
                SLAREKLSISNLDVKGLTSGLYLTYEQDAPLVLISQNTGWIFKDLYALLYHHLQLSDG
                HDDN"
gene            9683. .12172
                /gene="ORF08"
CDS             9683. .12172
                /gene="ORF08"
                /codon_start=1
                /product="ORF08"
                /protein_id="AAF59985.1"
                /db_xref="GI:7329996"
                /translation="MMITNRTRRLLRAWVVIIAIGTAVGENVTTPKGATTTAKPTPGP
                STPTPPENPPRAEAFKFRVCSASATGELFRFNLEKTCPGTEDKTHQEGILMVFKKNIV
                PHIFKVRRYRKVATSVTVYRGWTETAVTGKQEVIRPVPQYEINHMDTTYQCFSSMRVN
                VNGIVNTYTDRDFTNQTVFLQPVEGLTDNIQRYFSQPVLYTTPGWFPGIYRVRTTVNC
                EIVDMIARSAEPYSYFVTALGDTVEVSPFCHNDSTCSVAEKTENGLGVRALTNYTIVD
                FATRKPTTETRVFADSGEYTVSWKAEDPKSAVCALTLWKTFPRAIQTTHEASYHFVAN
                DVTATFTSPLSEVANFTGTYSCLDEVIQKTLNDTIKKLSDTHVTNGSAQYYKTEGGLF
                LLWQPLTPLSLVDEMRGLNGTTPAPPATTSTVSRVRRSVNTNEQATDNLAAPQLQFAY
                DKLRASINKVLEELSRAWCREQVRDTYMWYELSKINPTSVMTAIYGRPVSAKFVGDAI
                SVTDCVAVDQASVSIHKSLRTSTPGMCYSRPPVTFRFLNSTTLFKGQLGPRNEIILTD
                NQVEACKETCEHYFIASNVTYYYKDYVFVKKINTSEISTLGTFIALNLSFIENIDFRV
                IELYSRAEKKLSGSVFDIETMFREYNYYTQRLAGLREDLDNTIDLNRDRLARDLSEIV
                ADLGDVGRTVVNVASSVITLFGSIVSGFINFIKSPFGGMLMILVIVAVVLIVFALNRR
                TNAIAQAPIRMIYPDIDKMQPSGGKVDQEQIKNILAGMHQLQQEERRRLDEQQRSAPS
                LFRRASDGLKRRFRGYKPLENEEAQEYEMSK"
gene            12290. .15334
                /gene="ORF09"
CDS             12290. .15334
                /gene="ORF09"
                /codon_start=1
                /product="polymerase"
                /protein_id="AAF59986.1"
                /db_xref="GI:7329997"
                /translation="MDFFNPYLGPRGPRPHSHRGTDAPAPAGAGAVQPPPDVCRLIPA
                CLRTPGAGGMIPVTIPFPPTYFENGARGDVLLANERSMWTARDRKPVAPDPQDQSITF
                HAYDVVETTYAADRCAEVPSRFQTDIIPSGTVLKLLGRTEDGTSVCVNVFRQQVYFYA
                KVPAGINVTHILQQALKNTAGRAACGFSTRRVNKRILKTYDVAEHPVTEITLSSGSML
                STLSDRLVACGCEVFESNVDAVRRFVLDHGFTTFGWYSCARATPRLAARDARTALEFD
                CSWEDLSVQADRSDWPPYRIVAFDIECTGEAGFPCATRDGDAVIQISCVFYTTREGAP
                NPPNILFSVGTCDPIPDTDVLEFPSEYDMLVSFFAMIRDFEVDFLTGYNISNFDLPYL
                ITRASQVYNLRLNEYTKIKTGSIFEVHEPRGGGGGFMRSVSKIKIAGIVPIDMYQVCR
                EKLSLSDYKLDTVARQCLGGKKEDVSYKDIPPLFRSGPGGRAKVGSYCVMDSVLVMDL
                LKMFMIHVEISEIAKLAKIQARRVLTDGQQLRVFSCLLEAAARENFILPVPTPEGQGG
                YQGATVINPIPGFYDEPVLVVDFASLYPSIIQAHNLCYSTMIHGRDLHLHPNLTPDDY
                ETFVLSGGPVHFVKKHKRESLLGRLLTVWLEKRRAIRRTLAACDDPSLKTILDKQQLA
                IKVTCNAVYGFTGVASGLLPCINIAETVTLRGRTMLEMSKSYVEALTTEDLRTRLGRE
                VTARHGARFRVVYGDTDSLFIACDGYSAEAVSAFCDDLAARITADLFPPPIKLEAEKT
                FKCLLLLTKKRYIGVLLNDKMVMKGVDLIRKTACKFVQERCRAILDLVLHDPEVKAAA
                RLLCKRPPHAVYEEGLPAGFIKIVEVLNASYLDLRNSVVPIEQLTFSTELSRPVCDYK
                TTNLPHLAVYQKLASRCEELPQVHDRIPYVFVDAPGSLKSDLAEHPDYVRQHQIPVAV
                DLYFDKLVHGAANILQCLFGNNADTTVAILYNFLNVPYKLFS"
gene            15429. .16679
                /gene="ORF10"
CDS             15429. .16679
                /gene="ORF10"
                /codon_start=1
                /product="ORF10"
                /protein_id="AAF59987.1"
                /db_xref="GI:7329998"
                /translation="MLVNELSVVLGDWEVTFHRGRFSFVNLTRLQTFKGHGGYARVRL
                PFSLDQLLHQHFAFGLVTRLKELPPFSDCVALIAPLDSGGDADAARVAPGFVLDSSRP
                LTVWVNASGRHTIRFCLLFLKPIDLERAVTYVFGENGGARSEGTPKPTCATESLPGGP
                LRVSGEASQTSPHSFVAYFPTANSVACLSLLRLQVRPFSDDAAHRDARISPKYVTFSN
                SGGNVCKASVHTLSPSRCKTAQMEIIYAPGDPNAEIVLGQSGPVLPTHTGGRVLGVYA
                DAEKTIQPGSSAEVRVQLIFQQGAAARGDLAFLVTGVAPEPLFVVTPALLLSGCTTHL
                RLFNPNGTPTTIKRDTLVAAAAPCPVVRLSSADDAPRDLVASPDTGALSINAFTIPVG
                FPGVVSAECHVSLRDNGVHERMNH"
gene            16688. .17917
                /gene="ORF11"
CDS             16688. .17917
                /gene="ORF11"
                /codon_start=1
                /product="ORF11"
                /protein_id="AAF59988.1"
                /db_xref="GI:7329999"
                /translation="MGTPVRFFRGEWQTSSLVDNGTPRYSSLVWAATIHDGYLTLVNR
                SELCVTERSPCLPACPSIGRLVGKRFPGFAFASATLGDRGTRTVFYAFGHRDNPLDIV
                PAVVERADRELVLRVHAPQTTRVSRYGLKVFVAIVTVVRPPGVFLHFPQDRVPIALTD
```

Query Match 42.1%; Score 3823; DB 14; Length 130733;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 4073; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy       4996 GTAACACCATGTTAATAAAATACGCAGCATGCACATTTTTAGCTAAGCTCCTCAAAGACA 5055
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63947 GTAACACCATGTTAATAAAATACGCAGCATGCACATTTTTAGCTAAGCTCCTCAAAGACA 64006

Qy       5056 ATTTCCTCTTCAATTGATGACTCATCAACCTCGGTTTCCGTGTTAATGTCGGAAATATAT 5115
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64007 ATTTCCTCTTCAATTGATGACTCATCAACCTCGGTTTCCGTGTTAATGTCGGAAATATAT 64066

Qy       5116 GAGTCCAGGATACCCTCGTCATCGCCACAAATTAATTCCAGGGCTTGTAAAATATCATCC 5175
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64067 GAGTCCAGGATACCCTCGTCATCGCCACAAATTAATTCCAGGGCTTGTAAAATATCATCC 64126

Qy       5176 AGTGAATCAGCTGCTAATGTCAAGGATGTTGTTTCTAATTTTAGCCCATCCAAATGGTGA 5235
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64127 AGTGAATCAGCTGCTAATGTCAAGGATGTTGTTTCTAATTTTAGCCCATCCAAATGGTGA 64186

Qy       5236 GTGGGTGACACTTTGGGACAAGACCCCGTAAAATTGGCACGCTCCACACAAACGGAAGGA 5295
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64187 GTGGGCGACACTTTGGGACAAGACCCCGTAAAATTGGCACGCTCCACACAAACGGAAGGA 64246

Qy       5296 TGCATTTGATCGTTCACGAGGGGCAGCATATTTATTTTTATAACATCAGGAACCGGACCC 5355
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64247 TGCATTTGATCGTTCACGAGGGGCAGCATATTTATTTTTATAACATCAGGAACCGGACCC 64306

Qy       5356 GTCAATGACGCACATATGTCCATTAAGATTGGAGTCGTGACACTGGTTGAATTGACGGTG 5415
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64307 GTCAATGACGCACATATGTCCATTAAGATTGGAGTCGTGACACTGGTTGAATTGACGGTG 64366

Qy       5416 AAAAAAAATGTATTCGTGTCACATGCGTAACTATTTTTTAACCACACGAGGGCGGAATAA 5475
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64367 AAAAAAAATGTATTCGTGTCACATGCGTAACTATTTTTTAACCACACGAGGGCGGAATAA 64426

Qy       5476 AACGGGTCCCCGGCGCTAAGGCCGTGAATGCGGCAGGCACGGCTCATTGCCATTTCCAGA 5535
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64427 AACGGGTCCCCGGCGCTAAGGCCGTGAATGCGGCAGGCACGGCTCATTGCCATTTCCAGA 64486

Qy       5536 TTTGTTAAATCTGTTCGAACATACGAGTCTAAAATCAAGTATAACCACGACAAGTTCAAA 5595
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64487 TTTGTTAAATCTGTTCGAACATACGAGTCTAAAATCAAGTATAACCACGACAAGTTCAAA 64546

Qy       5596 CAGGAGGCAACGCGGCTGGTTTCCACCCGGACATCCCCGCTCACAAACCCGAACTCTCTA 5655
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64547 CAGGAGGCAACGCGGCTGGTTTCCACCCGGACATCCCCGCTCACAAACCCGAACTCTCTA 64606

Qy       5656 GGAACGCACATCAAAGCGTTCAGACAAAATTCCGAAACCGAAACCCTAGTTCTGAGCACC 5715
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db      64607 GGAACGCACATCAAAGCGTTCAAACAAAATTCCGAAACCGAAACCCTAGTTCTGAGCACC 64666

Qy       5716 ATCTGGTTATAAACTGCGGTTAAGAGACGAGCGGCCACACACTGCTGCTTGTGAGGATTC 5775
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64667 ATCTGGTTATAAACTGCGGTTAAGAGACGAGCGGCCACACACTGCTGCTTGTGAGGATTC 64726
```

```
Qy       5776 AGCTTAGGTGGCCTGCAGGCTTGCTGACAGGCCCGTAGCTGCCTGGCGGCGCTCGCGCAC 5835
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64727 AGCTTAGGTGGCCTGCAGGCTTGCTGACAGGCCCGTAGCTGCCTGGCGGCGCTCGCGCAC 64786

Qy       5836 CTTTGACAGCCGCACGCCAGTTCCAGAGCGAGGCAGAAGCGCTCTTTGCAGTCGCGCCAT 5895
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64787 CTTTGACAGCCGCACGCCAGTTCCAGAGCGAGGCAGAAGCGCTCTTTGCAGTCGCGCCAT 64846

Qy       5896 ATCTCGGCCATAGGAGATTCGCTCGCGTGATCTTTTAGGTGCATATATATCTGTTGGGTA 5955
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64847 ATCTCGGCCATAGGAGATTCGCTCGCGTGATCTTTTAGGTGCATATATATCTGTTGGGTA 64906

Qy       5956 ACAAACCTCACTCCCTGTAAAAGGGGAATAAGGTCCGTTCTTTCTATCGTTTTCTCTGAA 6015
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64907 ACAAACCTCACTCCCTGTAAAAGGGGAATAAGGTCCGTTCTTTCTATCGTTTTCTCTGAA 64966

Qy       6016 TCTATAGTTACATCTTTAAGTATAACCAAAGACGCTAAAAACCCAGGATGGTCGACACCT 6075
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64967 TCTATAGTTACATCTTTAAGTATAACCAAAGACGCTAAAAACCCAGGATGGTCGACACCT 65026

Qy       6076 TTAAAAAATCGCAAAAGCGACCGGATGGCTCTATCCGCGTTTCCATGTCGTTCAAAGGTA 6135
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65027 TTAAAAAATCGCAAAAGCGACCGGATGGCTCTATCCGCGTTTCCATGTCGTTCAAAGGTA 65086

Qy       6136 GTCACGATGGATCTCCAGTTAGACTCTGTCTCCCTGTTAACACCCTTAACTGGAATGGAA 6195
             |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      65087 GTCACGATGGATCTCCAGTTAGACTCGGTCTCCCTGTTAACACCCTTAACTGGAATGGAA 65146

Qy       6196 ACAGCCATGACACCGTGAACTTCCTGATGTCTCCTAAAACTAACCCCGAAACAGAGCTAA 6255
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65147 ACAGCCATGACACCGTGAACTTCCTGATGTCTCCTAAAACTAACCCCGAAACAGAGCTAA 65206

Qy       6256 ATACCAATGACTGTCACCCCTACCCCAAGCCACGCCCCCGTACTATTAGACCAGGGTGAG 6315
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65207 ATACCAATGACTGTCACCCCTACCCCAAGCCACGCCCCCGTACTATTAGACCAGGGTGAG 65266

Qy       6316 TAACCACGCTATCCTTTAAAAACCCATACGTGGAGTTTGTAAGGTAAACAGCTCGTGTAT 6375
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65267 TAACCACGCTATCCTTTAAAAACCCATACGTGGAGTTTGTAAGGTAAACAGCTCGTGTAT 65326

Qy       6376 ATTTCAGACGCCTGTCATAAAATGGATACCGACGACAATCAGGTAATTAAACTTTTTTTA 6435
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65327 ATTTCAGACGCCTGTCATAAAATGGATACCGACGACAATCAGGTAATTAAACTTTTTTTA 65386

Qy       6436 TTCAAGATTCAGGAGGGCGTGTTTACAGTGGTGTAGGTGGGAGCATATCTCGCCAATGGG 6495
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65387 TTCAAGATTCAGGAGGGCGTGTTTACAGTGGTGTAGGTGGGAGCATATCTCGCCAATGGG 65446

Qy       6496 AATGGCTGATGAATCCACACTTAGTGCTCGGATCAGGTGCTCTGTGTACTTTACTAGTGC 6555
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65447 AATGGCTGATGAATCCACACTTAGTGCTCGGATCAGGTGCTCTGTGTACTTTACTAGTGC 65506

Qy       6556 CGATGAACCTAAATTTGGCGTGTTTAGGTCGTACCAACCAAAAAACTGCCTGGCCTCGCC 6615
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65507 CGATGAACCTAAATTTGGCGTGTTTAGGTCGTACCAACCAAAAAACTGCCTGGCCTCGCC 65566

Qy       6616 TTTTAGAAGCTCTATGCAAGCTTTAATCGTGTCCAATAGCTTGTCTTTAAAAATGCAACT 6675
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65567 TTTTAGAAGCTCTATGCAAGCTTTAATCGTGTCCAATAGCTTGTCTTTAAAAATGCAACT 65626

Qy       6676 CTGGTACATCTTTATGACAGTGGTCCAAAAAAAACAAAGATTTAAAAACACGTTAAACTG 6735
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65627 CTGGTACATCTTTATGACAGTGGTCCAAAAAAAACAAAGATTTAAAAACACGTTAAACTG 65686

Qy       6736 CGTGTCCTGGTAGTCCTCGTATATAACCTCCTCAACAAGAAAAAATTTTTTAACCAAACT 6795
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65687 CGTGTCCTGGTAGTCCTCGTATATAACCTCCTCAACAAGAAAAAATTTTTTAACCAAACT 65746

Qy       6796 CGCCAGGTACTGAAACGATGGTACGGACAGGTCGTGAAAAGTGTCTATCATCTCCTTCAC 6855
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65747 CGCCAGGTACTGAAACGATGGTACGGACAGGTCGTGAAAAGTGTCTATCATCTCCTTCAC 65806

Qy       6856 CTCCCCTTTGAAGGTTTTGGTTACACCGACCATGTGAGATAGGCACCAGCTCAGGGGGGA 6915
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65807 CTCCCCTTTGAAGGTTTTGGTTACACCGACCATGTGAGATAGGCACCAGCTCAGGGGGGA 65866

Qy       6916 GGTCGGATCGTGAGGGGGTAACAATTCGTGGGCGTGGGGGTACTCCAGTTTTAGGATCTC 6975
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65867 GGTCGGATCGTGAGGGGGTAACAATTCGTGGGCGTGGGGGTACTCCAGTTTTAGGATCTC 65926

Qy       6976 CGGCAGGACGCGTACGAGTTCTTCGTCAAATCCTACGCAGCCCGCATCCACCAGGGGCAG 7035
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65927 CGGCAGGACGCGTACGAGTTCTTCGTCAAATCCTACGCAGCCCGCATCCACCAGGGGCAG 65986

Qy       7036 TACTGAGTTAATAAGTCCCCGAATTCTATCGTTCGTAATTTTGTGGAGCTCCTTTAGACA 7095
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65987 TACTGAGTTAATAAGTCCCCGAATTCTATCGTTCGTAATTTTGTGGAGCTCCTTTAGACA 66046

Qy       7096 GTAAACCATGTTGCTCCGACACTGCGGTACGATGAACCGCTCTCCTTGTCTGTGTTGCAT 7155
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66047 GTAAACCATGTTGCTCCGACACTGCGGTACGATGAACCGCTCTCCTTGTCTGTGTTGCAT 66106

Qy       7156 GAGCACCTGGAGTATCTTAGACAGGTATAGGGCCCGTTCCACTCTAAGGGTTGCCGCTTG 7215
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66107 GAGCACCTGGAGTATCTTAGACAGGTATAGGGCCCGTTCCACTCTAAGGGTTGCCGCTTG 66166

Qy       7216 GTAGAGCGGATTGCGGGCATTGACGGTCAGTGTGTTCATTTCACCGACTATCATGCCGAG 7275
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66167 GTAGAGCGGATTGCGGGCATTGACGGTCAGTGTGTTCATTTCACCGACTATCATGCCGAG 66226

Qy       7276 TGGCGGATAGTGGAACTTGTACATATGATTTAAAAGATGGTCTTTCCCGTAATGGCGACT 7335
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66227 TGGCGGATAGTGGAACTTGTACATATGATTTAAAAGATGGTCTTTCCCGTAATGGCGACT 66286

Qy       7336 CATGGTCGTGTTTTAACGTTCACAGGGTGTTAAAAAACTCAGAAGGTCCTGCGTGGACAA 7395
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66287 CATGGTCGTGTTTTAACGTTCACAGGGTGTTAAAAAACTCAGAAGGTCCTGCGTGGACAA 66346

Qy       7396 CTTTTGCGGACTCTCCGAACCGCTCAGGGATAAGCTGTACCACGCGGTGCTTCTTATAAA 7455
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66347 CTTTTGCGGACTCTCCGAACCGCTCAGGGATAAGCTGTACCACGCGGTGCTTCTTATAAA 66406

Qy       7456 CGAAGCATACGTCAAAAAACATGACCCACGTGACCTAAATAAGCTCACAGAAGAAGTCTG 7515
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66407 CGAAGCATACGTCAAAAAACATGACCCACGTGACCTAAATAAGCTCACAGAAGAAGTCTG 66466

Qy       7516 TCGATGTATTGTAATGGAATGTGCCTCCCTGGGTCCCATCTCCGGGCTCATCGCGGATCT 7575
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66467 TCGATGTATTGTAATGGAATGTGCCTCCCTGGGTCCCATCTCCGGGCTCATCGCGGATCT 66526

Qy       7576 GAATCTCTTTAACCTGTTCTGCCTCTACCGGGGTTCCCGGGTTAAAACCCGCGGGGCCGC 7635
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66527 GAATCTCTTTAACCTGTTCTGCCTCTACCGGGGTTCCCGGGTTAAAACCCGCGGGGCCGC 66586

Qy       7636 CACCTGTAACGTCCCGTGCGCAGAATGCGCCCAGGGCATCGTGAGAATTCTCACCGAGAG 7695
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66587 CACCTGTAACGTCCCGTGCGCAGAATGCGCCCAGGGCATCGTGAGAATTCTCACCGAGAG 66646

Qy       7696 GGCCCTTTGTTGTACCGAAAAGATGTTCATAGCGTCTGCCTGCAGCGGGGTCGTTATACC 7755
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66647 GGCCCTTTGTTGTACCGAAAAGATGTTCATAGCGTCTGCCTGCAGCGGGGTCGTTATACC 66706

Qy       7756 GCCCCAGTTGGCCAGAGTTCTGCACGACGTCTACGCGGAAATGAAGGCCAAGTGTTTGGG 7815
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66707 GCCCCAGTTGGCCAGAGTTCTGCACGACGTCTACGCGGAAATGAAGGCCAAGTGTTTGGG 66766

Qy       7816 GGCGTGGCGTCGTCTCATATGTTGCAGGCGGCCTATCATGGCTATTGCCGACTCGGTCCT 7875
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66767 GGCGTGGCGTCGTCTCATATGTTGCAGGCGGCCTATCATGGCTATTGCCGACTCGGTCCT 66826

Qy       7876 CGTCACTTATAATACCCTGGATGCCGAGGGAAAACTGGAACTTAGACTCAAAGCACTGTG 7935
             |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db      66827 CGTCACTTATAATACCCTGGATGCCGAGGGAAAACTGGACCTTAGACTCAAAGCACTGTG 66886

Qy       7936 CAAACTTGTTTTTCAACCCATCTTTCTTCAAAGAATCTTAGCCCCTATGCAGCTACTGGC 7995
```

```
Db      66887 CAAACTTGTTTTTCAACCCATCTTTCTGCAAAGAATCTTAGCCCCTATGCAGCTACTGGC 66946

Qy       7996 CAACGGGAAGATGGTTCCTGACAATTATTTTACCATCACCGGTACGGCCGAGAAGAGGCG 8055
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66947 CAACGGGAAGATGGTTCCTGACAATTATTTTACCATCACCGGTACGGCCGAGAAGAGGCG 67006

Qy       8056 CCCTGTCGTGACTGGAAGTACTAGCGGGATGACGTGTCCGGGAAGCAGCCTTGTCCCCGA 8115
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67007 CCCTGTCGTGACTGGAAGTACTAGCGGGATGACGTGTCCGGGAAGCAGCCTTGTCCCCGA 67066

Qy       8116 TTCCTTAATCCTGCCAGTATGCGAACCGGGGCTTCTCCCGGCACCCCTGGTTGACCTCAG 8175
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67067 TTCCTTAATCCTGCCAGTATGCGAACCGGGGCTTCTCCCGGCACCCCTGGTTGACCTCAG 67126

Qy       8176 TAATGTCTTAGAAAATCCAGAAATCATCCTCAGCGCCCCACCCCTGAGTCAATTTGTCAT 8235
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67127 TAATGTCTTAGAAAATCCAGAAATCATCCTCAGCGCCCCACCCCTGAGTCAATTTGTCAT 67186

Qy       8236 CACAAACACGCACCCCAGTCTGCCTCAGTCAGTCAGCATTATTACGCCAACCCAGGGCGT 8295
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67187 CACAAACACGCACCCCAGTCTGCCTCAGTCAGTCAGCATTATTACGCCAACCCAGGGCGT 67246

Qy       8296 TGTTCCCGGCCAATGTTTTATGGACACGTGGAAAGCGGTGTCACAGAGCATTCACCACCA 8355
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67247 TGTTCCCGGCCAATGTTTTATGGACACGTGGAAAGCGGTGTCACAGAGCATTCACCACCA 67306

Qy       8356 GGCACAGACGCCTATTTTGGCCGCCGCACTAACCGGTTCGACATCTGCGGCCCCTGGCCC 8415
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67307 GGCACAGACGCCTATTTTGGCCGCCGCACTAACCGGTTCGACATCTGCGGCCCCTGGCCC 67366

Qy       8416 GCATATCGCATGTTCCCCAGTTGCCGGCACGTCTCGGCAGGTGGAAGGGTCCGCGGGCGT 8475
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67367 GCATATCGCATGTTCCCCAGTTGCCGGCACGTCTCGGCAGGTGGAAGGGTCCGCGGGCGT 67426

Qy       8476 CGATTGCGGGAAACCAGCATGCGTTCCGCAGCCCGCGTTACCGCCCAATGTCCCCGCCAA 8535
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67427 CGATTGCGGGAAACCAGCATGCGTTCCGCAGCCCGCGTTACCGCCCAATGTCCCCGCCAA 67486

Qy       8536 GAGGATGGAAACGGTAGCACAGTTGGGAAACGCTCCCGTAAAAAACGTCCACATCGGAGG 8595
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67487 GAGGATGGAAACGGTAGCACAGTTGGGAAACGCTCCCGTAAAAAACGTCCACATCGGAGG 67546

Qy       8596 CCGCGTATACGCTCCACTGGTTAATATACCAATAATAGACTTAACGTCCCCGTCAGGGTC 8655
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67547 CCGCGTATACGCTCCACTGGTTAATATACCAATAATAGACTTAACGTCCCCGTCAGGGTC 67606

Qy       8656 CGGCCAGAGTCCGGCCGATATCGCCAACACTCCAGAGTCCCGCATGGCGGCCGGCTCTCC 8715
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67607 CGGCCAGAGTCCGGCCGATATCGCCAACACTCCAGAGTCCCGCATGGCGGCCGGCTCTCC 67666

Qy       8716 GCCCTTCGCCGAAACCGCCGCAACGGTCCCCGCTAAGAGAAAGCAGCCACGCGAGGACGT 8775
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67667 GCCCTTCGCCGAAACCGCCGCAACGGTCCCCGCTAAGAGAAAGCAGCCACGCGAGGACGT 67726

Qy       8776 GGCAGACAAAAGACTGAAGGGCGACGTTCGGGGCGCCGCAACAGTAAACCACCCTTTCCC 8835
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67727 GGCAGACAAAAGACTGAAGGGCGACGTTCGGGGCGCCGCAACAGTAAACCACCCTTTCCC 67786

Qy       8836 GGGACCGTCCGGGATGCGCGTTCGCGAGCAGGGCTTATTCGATTTAATCGAAAGCTCCAC 8895
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67787 GGGACCGTCCGGGATGCGCGTTCGCGAGCAGGGCTTATTCGATTTAATCGAAAGCTCCAC 67846

Qy       8896 GGATGTAACCGCGAACGCATCTGGACCCAAAAACGACGACGACATGCTAGCGGCTATCCT 8955
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67847 GGATGTAACCGCGAACGCATCTGGACCCAAAAACGACGACGACATGCTAGCGGCTATCCT 67906

Qy       8956 ACAGGACCTGTATGGACTACAGTCCCCCCCGGCCATCGATTCCCCCTCCAGCAACTCGGA 9015
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67907 ACAGGACCTGTATGGACTACAGTCCCCCCCGGCCATCGATTCCCCCTCCAGCAACTCGGA 67966

Qy       9016 CAATGAGGAGATATTTCCAGAGGTTAGTCCGCCATCTAGCGGCCACGGATCGCCTTGA 9073
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      67967 CAATGAGGAGATATTTCCAGAGGTTAGTCCGCCATCTAGCGGCCACGGATCGCCTTGA 68024
```

RESULT 3
AF241163

```
LOCUS       AF241163               1945 bp    mRNA    linear   VRL 23-OCT-2002
DEFINITION  Cercopithecine herpesvirus 17 immediate-early transcription factor
            RRta mRNA, complete cds.
ACCESSION   AF241163
VERSION     AF241163.1  GI:13506778
KEYWORDS    .
SOURCE      Cercopithecine herpesvirus 17
  ORGANISM  Cercopithecine herpesvirus 17
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Rhadinovirus.
REFERENCE   1  (bases 1 to 1945)
  AUTHORS   Lin,S.F., Robinson,D.R., Oh,J., Jung,J.U., Luciw,P.A. and Kung,H.J.
  TITLE     Identification of the bZIP and Rta homologues in the genome of
            rhesus monkey rhadinovirus
  JOURNAL   Virology 298 (2), 181-188 (2002)
  MEDLINE   22127395
   PUBMED   12127781
REFERENCE   2  (bases 1 to 1945)
  AUTHORS   Lin,S.F., Robinson,D.R., Oh,J., Kung,H.J. and Luciw,P.A.
  TITLE     Cloning and characterization of cDNAs encoding Rhesus Rhadinovirus
            (RRV) ORF50
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1945)
  AUTHORS   Lin,S.F., Robinson,D.R., Oh,J., Kung,H.J. and Luciw,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAR-2000) Biological Chemistry, School of Medicine,
            UC Davis, 4645 Second Avenue, Rm. 2400, Sacramento, CA 95831, USA
FEATURES             Location/Qualifiers
     source          1. .1945
                     /organism="Cercopithecine herpesvirus 17"
                     /mol_type="mRNA"
                     /isolate="17577"
                     /specific_host="Macaca mulatta"
                     /db_xref="taxon:154334"
     CDS             65. .1798
                     /note="ORF50"
                     /codon_start=1
                     /product="immediate-early transcription factor RRta"
                     /protein_id="AAK28330.1"
                     /db_xref="GI:13506779"
                     /translation="MDTDDNQGVKKLRRSCVDNFCGLSEPLRDKLYHAVLLINEAYVK
                     KHDPRDLNKLTEEVCRCIVMECASLGPISGLIADLNLFNLFCLYRGSRVKTRGAATCN
                     VPCAECAQGIVRILTERALCCTEKMFIASACSGVVIPPQLARVLHDVYAEMKAKCLGA
                     WRRLICCRRPIMAIADSVLVTYNTLDAEGKLDLRLKALCKLVFQPIFLQRILAPMQLL
                     ANGKMVPDNYFTITGTAEKRRPVVTGSTSGMTCPGSSLVPDSLILPVCEPGLLPAPLV
                     DLSNVLENPEIILSAPPLSQFVITNTHPSLPQSVSIITPTQGVVPGQCFMDTWKAVSQ
                     SIHHQAQTPILAAALTGSTSAAPGPHIACSPVAGTSRQVEGSAGVDCGKPACVPQPAL
                     PPNVPAKRMETVAQLGNAPVKNVHIGGRVYAPLVNIPIIDLTSPSGSGQSPADIANTP
                     ESRMAAGSPPFAETAATVPAKRKQPREDVADKRLKGDVRGAATVNHPFPGPSGMRVRE
                     QGLFDLIESSTDVTANASGPKNDDDMLAAILQDLYGLQSPPAIDSPSSNSDNEEIFPE
                     VSPPSSGHGSP"
BASE COUNT      471 a    581 c    490 g    403 t
ORIGIN
```

```
  Query Match            18.4%;  Score 1665;  DB 14;  Length 1945;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1715;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       7358 CAGGGTGTTAAAAAACTCAGAAGGTCCTGCGTGGACAACTTTTGCGGACTCTCCGAACCG 7417
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 CAGGGTGTTAAAAAACTCAGAAGGTCCTGCGTGGACAACTTTTGCGGACTCTCCGAACCG 142

Qy       7418 CTCAGGGATAAGCTGTACCACGCGGTGCTTCTTATAAACGAAGCATACGTCAAAAAACAT 7477
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 CTCAGGGATAAGCTGTACCACGCGGTGCTTCTTATAAACGAAGCATACGTCAAAAAACAT 202

Qy       7478 GACCCACGTGACCTAAATAAGCTCACAGAAGAAGTCTGTCGATGTATTGTAATGGAATGT 7537
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        '203 GACCCACGTGACCTAAATAAGCTCACAGAAGAAGTCTGTCGATGTATTGTAATGGAATGT 262

Qy       7538 GCCTCCCTGGGTCCCATCTCCGGGCTCATCGCGGATCTGAATCTCTTTAACCTGTTCTGC 7597
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 GCCTCCCTGGGTCCCATCTCCGGGCTCATCGCGGATCTGAATCTCTTTAACCTGTTCTGC 322

Qy       7598 CTCTACCGGGGTTCCCGGGTTAAAACCCGCGGGGCCGCCACCTGTAACGTCCCGTGCGCA 7657
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 CTCTACCGGGGTTCCCGGGTTAAAACCCGCGGGGCCGCCACCTGTAACGTCCCGTGCGCA 382

Qy       7658 GAATGCGCCCAGGGCATCGTGAGAATTCTCACCGAGAGGGCCCTTTGTTGTACCGAAAAG 7717
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GAATGCGCCCAGGGCATCGTGAGAATTCTCACCGAGAGGGCCCTTTGTTGTACCGAAAAG 442

Qy       7718 ATGTTCATAGCGTCTGCCTGCAGCGGGGTCGTTATACCGCCCCAGTTGGCCAGAGTTCTG 7777
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 ATGTTCATAGCGTCTGCCTGCAGCGGGGTCGTTATACCGCCCCAGTTGGCCAGAGTTCTG 502

Qy       7778 CACGACGTCTACGCGGAAATGAAGGCCAAGTGTTTGGGGGCGTGGCGTCGTCTCATATGT 7837
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 CACGACGTCTACGCGGAAATGAAGGCCAAGTGTTTGGGGGCGTGGCGTCGTCTCATATGT 562

Qy       7838 TGCAGGCGGCCTATCATGGCTATTGCCGACTCGGTCCTCGTCACTTATAATACCCTGGAT 7897
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 TGCAGGCGGCCTATCATGGCTATTGCCGACTCGGTCCTCGTCACTTATAATACCCTGGAT 622

Qy       7898 GCCGAGGGAAAACTGGAACTTAGACTCAAAGCACTGTGCAAACTTGTTTTTCAACCCATC 7957
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        623 GCCGAGGGAAAACTGGACCTTAGACTCAAAGCACTGTGCAAACTTGTTTTTCAACCCATC 682

Qy       7958 TTTCTTCAAAGAATCTTAGCCCCTATGCAGCTACTGGCCAACGGGAAGATGGTTCCTGAC 8017
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        683 TTTCTTCAAAGAATCTTAGCCCCTATGCAGCTACTGGCCAACGGGAAGATGGTTCCTGAC 742

Qy       8018 AATTATTTTACCATCACCGGTACGGCCGAGAAGAGGCGCCCTGTCGTGACTGGAAGTACT 8077
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 AATTATTTTACCATCACCGGTACGGCCGAGAAGAGGCGCCCTGTCGTGACTGGAAGTACT 802

Qy       8078 AGCGGGATGACGTGTCCGGGAAGCAGCCTTGTCCCCGATTCCTTAATCCTGCCAGTATGC 8137
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 AGCGGGATGACGTGTCCGGGAAGCAGCCTTGTCCCCGATTCCTTAATCCTGCCAGTATGC 862

Qy       8138 GAACCGGGGCTTCTCCCGGCACCCCTGGTTGACCTCAGTAATGTCTTAGAAAATCCAGAA 8197
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        863 GAACCGGGGCTTCTCCCGGCACCCCTGGTTGACCTCAGTAATGTCTTAGAAAATCCAGAA 922

Qy       8198 ATCATCCTCAGCGCCCCACCCCTGAGTCAATTTGTCATCACAAACACGCACCCCAGTCTG 8257
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        923 ATCATCCTCAGCGCCCCACCCCTGAGTCAATTTGTCATCACAAACACGCACCCCAGTCTG 982

Qy       8258 CCTCAGTCAGTCAGCATTATTACGCCAACCCAGGGCGTTGTTCCCGGCCAATGTTTTATG 8317
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 CCTCAGTCAGTCAGCATTATTACGCCAACCCAGGGCGTTGTTCCCGGCCAATGTTTTATG 1042

Qy       8318 GACACGTGGAAAGCGGTGTCACAGAGCATTCACCACCAGGCACAGACGCCTATTTTGGCC 8377
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1043 GACACGTGGAAAGCGGTGTCACAGAGCATTCACCACCAGGCACAGACGCCTATTTTGGCC 1102

Qy       8378 GCCGCACTAACCGGTTCGACATCTGCGGCCCCTGGCCCGCATATCGCATGTTCCCCAGTT 8437
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1103 GCCGCACTAACCGGTTCGACATCTGCGGCCCCTGGCCCGCATATCGCATGTTCCCCAGTT 1162

Qy       8438 GCCGGCACGTCTCGGCAGGTGGAAGGGTCCGCGGGCGTCGATTGCGGGAAACCAGCATGC 8497
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1163 GCCGGCACGTCTCGGCAGGTGGAAGGGTCCGCGGGCGTCGATTGCGGGAAACCAGCATGC 1222

Qy       8498 GTTCCGCAGCCCGCGTTACCGCCCAATGTCCCCGCCAAGAGGATGGAAACGGTAGCACAG 8557
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1223 GTTCCGCAGCCCGCGTTACCGCCCAATGTCCCCGCCAAGAGGATGGAAACGGTAGCACAG 1282

Qy       8558 TTGGGAAACGCTCCCGTAAAAAACGTCCACATCGGAGGCCGCGTATACGCTCCACTGGTT 8617
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1283 TTGGGAAACGCTCCCGTAAAAAACGTCCACATCGGAGGCCGCGTATACGCTCCACTGGTT 1342

Qy       8618 AATATACCAATAATAGACTTAACGTCCCCGTCAGGGTCCGGCCAGAGTCCGGCCGATATC 8677
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1343 AATATACCAATAATAGACTTAACGTCCCCGTCAGGGTCCGGCCAGAGTCCGGCCGATATC 1402

Qy       8678 GCCAACACTCCAGAGTCCCGCATGGCGGCCGGCTCTCCGCCCTTCGCCGAAACCGCCGCA 8737
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1403 GCCAACACTCCAGAGTCCCGCATGGCGGCCGGCTCTCCGCCCTTCGCCGAAACCGCCGCA 1462

Qy       8738 ACGGTCCCCGCTAAGAGAAAGCAGCCACGCGAGGACGTGGCAGACAAAAGACTGAAGGGC 8797
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1463 ACGGTCCCCGCTAAGAGAAAGCAGCCACGCGAGGACGTGGCAGACAAAAGACTGAAGGGC 1522

Qy       8798 GACGTTCGGGGCGCCGCAACAGTAAACCACCCTTTCCCGGGACCGTCCGGGATGCGCGTT 8857
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1523 GACGTTCGGGGCGCCGCAACAGTAAACCACCCTTTCCCGGGACCGTCCGGGATGCGCGTT 1582

Qy       8858 CGCGAGCAGGGCTTATTCGATTTAATCGAAAGCTCCACGGATGTAACCGCGAACGCATCT 8917
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1583 CGCGAGCAGGGCTTATTCGATTTAATCGAAAGCTCCACGGATGTAACCGCGAACGCATCT 1642

Qy       8918 GGACCCAAAAACGACGACGACATGCTAGCGGCTATCCTACAGGACCTGTATGGACTACAG 8977
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1643 GGACCCAAAAACGACGACGACATGCTAGCGGCTATCCTACAGGACCTGTATGGACTACAG 1702

Qy       8978 TCCCCCCCGGCCATCGATTCCCCCTCCAGCAACTCGGACAATGAGGAGATATTTCCAGAG 9037
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1703 TCCCCCCCGGCCATCGATTCCCCCTCCAGCAACTCGGACAATGAGGAGATATTTCCAGAG 1762

Qy       9038 GTTAGTCCGCCATCTAGCGGCCACGGATCGCCTTGA 9073
             ||||||||||||||||||||||||||||||||||||
Db       1763 GTTAGTCCGCCATCTAGCGGCCACGGATCGCCTTGA 1798
```

```
RESULT 4
AL139384/c
LOCUS       AL139384              187157 bp    DNA     linear   PRI 20-JUN-2001
DEFINITION  Human DNA sequence from clone RP11-88E10 on chromosome 13q33.1-34,
            complete sequence.
ACCESSION   AL139384
VERSION     AL139384.17  GI:14530206
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 187157)
  AUTHORS   Smith,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 22, 2001 this sequence version replaced gi:11610962.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-88E10 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence was finished as follows unless otherwise noted: all
```

regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. IMPORTANT: This sequence is not the entire insert of clone RP11-88E10 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RP11-88E10 is at 1 in this sequence. The true left end of clone RP11-480K16 is at 187058 in this sequence.
The true right end of clone RP11-144L2 is at 50570 in this sequence.

FEATURES	Location/Qualifiers
source	1. .187157
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="13"
	/map="q33.1-34"
	/clone="RP11-88E10"
	/clone_lib="RPCI-11.1"
repeat_region	1046. .1077
	/note="MER75 repeat: matches 333. .364 of consensus"
repeat_region	1494. .1517
	/note="12 copies 2 mer ac 100% conserved"
repeat_region	1520. .1818
	/note="AluSg repeat: matches 1. .298 of consensus"
misc_feature	2127. .3461
	/note="CpG island"
	/evidence=not_experimental
repeat_region	3035. .3126
	/note="46 copies 2 mer cc 64% conserved"
repeat_region	3136. .3177
	/note="6 copies 7 mer gggcaga 83% conserved"
misc_feature	complement(4293. .4718)
	/note="match: STS: Em:G29775"
repeat_region	4299. .4550
	/note="LTR33 repeat: matches 256. .516 of consensus"
repeat_region	5281. .5623
	/note="MLT1A1 repeat: matches 1. .365 of consensus"
repeat_region	5892. .6094
	/note="MLT1H repeat: matches 79. .291 of consensus"
repeat_region	7071. .7380
	/note="AluSx repeat: matches 1. .310 of consensus"
repeat_region	7620. .7668
	/note="MER5B repeat: matches 1. .49 of consensus"
repeat_region	7692. .7953
	/note="AluJo repeat: matches 1. .285 of consensus"
misc_feature	complement(7708. .8008)
	/note="match: GSS: Em:AQ757664"
repeat_region	7993. .8051
	/note="MER5B repeat: matches 40. .104 of consensus"
misc_feature	8068. .8448
	/note="match: GSS: Em:AQ014714"
misc_feature	complement(8224. .8765)
	/note="match: GSS: Em:AQ141652"
misc_feature	complement(8600. .8786)
	/note="match: STS: Em:G50946
	match: GSS: Em:B82990"
misc_feature	complement(8612. .8786)
	/note="match: GSS: Em:B83029"
repeat_region	9048. .9133
	/note="2 copies 43 mer 94% conserved"
repeat_region	9922. .10225
	/note="AluSg repeat: matches 1. .305 of consensus"
repeat_region	10390. .10705
	/note="L1MB3 repeat: matches 5870. .6182 of consensus"
repeat_region	10731. .11023
	/note="AluY repeat: matches 1. .293 of consensus"
repeat_region	11112. .11357
	/note="AluSg/x repeat: matches 70. .301 of consensus"
repeat_region	11359. .11644
	/note="AluJb repeat: matches 1. .286 of consensus"
repeat_region	13908. .14343
	/note="SVA repeat: matches 956. .1385 of consensus"
repeat_region	14344. .15086
	/note="SVA repeat: matches 6. .760 of consensus"
repeat_region	15147. .15443
	/note="AluSx repeat: matches 1. .295 of consensus"
repeat_region	15895. .16203
	/note="AluSq repeat: matches 1. .310 of consensus"
repeat_region	16519. .16878
	/note="HAL1 repeat: matches 721. .1114 of consensus"
misc_feature	complement(18798. .19280)
	/note="match: GSS: Em:AQ147060"
repeat_region	19105. .19481
	/note="MLT1E repeat: matches 176. .568 of consensus"
repeat_region	19485. .19784
	/note="AluSx repeat: matches 2. .302 of consensus"
misc_feature	complement(19635. .20146)
	/note="match: GSS: Em:AQ395327"
repeat_region	19822. .20067
	/note="6 copies 41 mer 65% conserved"
repeat_region	19855. .20082
	/note="12 copies 19 mer 64% conserved"
repeat_region	19857. .20072
	/note="8 copies 27 mer 63% conserved"
repeat_region	19911. .20082
	/note="4 copies 43 mer 70% conserved"
repeat_region	20932. .21232
	/note="AluSx repeat: matches 1. .302 of consensus"
misc_feature	complement(21053. .21546)
	/note="match: GSS: Em:AQ512552"
misc_feature	complement(21119. .21534)
	/note="match: GSS: Em:AQ564752"
repeat_region	21393. .21568
	/note="MIR repeat: matches 36. .223 of consensus"
misc_feature	complement(21587. .21880)
	/note="match: STS: Em:HSPH07E6"
repeat_region	21789. .21843
	/note="TAR1 repeat: matches 635. .686 of consensus"
repeat_region	22130. .22437
	/note="AluSx repeat: matches 1. .308 of consensus"
misc_feature	23147. .23486
	/note="match: GSS: Em:AQ080832"
repeat_region	23229. .23508
	/note="AluSx repeat: matches 1. .304 of consensus"
misc_feature	complement(23618. .24039)
	/note="match: GSS: Em:AQ470695"
repeat_region	24469. .24761
	/note="AluSx repeat: matches 1. .294 of consensus"
repeat_region	24884. .25757
	/note="L1M4 repeat: matches 3366. .4222 of consensus"
repeat_region	25960. .27013
	/note="527 copies 2 mer ac 69% conserved"
repeat_region	27026. .27191
	/note="83 copies 2 mer ca 72% conserved"
repeat_region	28259. .28394
	/note="MER5A repeat: matches 29. .189 of consensus"
repeat_region	28938. .29244
	/note="AluSq repeat: matches 1. .298 of consensus"
repeat_region	29249. .29548
	/note="AluSx repeat: matches 1. .302 of consensus"
repeat_region	29676. .30494
	/note="L1MA6 repeat: matches 4193. .5002 of consensus"
repeat_region	30495. .30795
	/note="AluY repeat: matches 2. .302 of consensus"
repeat_region	30796. .31938
	/note="L1MA6 repeat: matches 5002. .6255 of consensus"
repeat_region	31941. .32000
	/note="MER41B repeat: matches 483. .547 of consensus"
repeat_region	32003. .32059
	/note="L1MA5 repeat: matches 6242. .6300 of consensus"
repeat_region	32103. .32448
	/note="TIGGER2 repeat: matches 1. .357 of consensus"

```
     repeat_region   32449. .32758
                     /note="AluSx repeat: matches 1. .311 of consensus"
     repeat_region   32759. .32822
                     /note="TIGGER2 repeat: matches 357. .422 of consensus"
     repeat_region   32823. .33132
                     /note="AluY repeat: matches 2. .310 of consensus"
     repeat_region   33133. .33489
                     /note="TIGGER2 repeat: matches 422. .744 of consensus"
     repeat_region   33553. .33764
                     /note="TIGGER2 repeat: matches 2462. .2690 of consensus"
     repeat_region   34943. .35023
                     /note="MER81 repeat: matches 7. .88 of consensus"
     misc_feature    complement(35414. .35877)
                     /note="match: GSS: Em:AQ220607"
     misc_feature    complement(35464. .35877)
                     /note="match: GSS: Em:AQ451480"
     repeat_region   35690. .35727

  Query Match             0.3%;  Score 25;  DB 9;  Length 187157;
  Best Local Similarity   100.0%;  Pred. No. 0.44;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4152 GTTTTAAAAATGGAGAAAGATTTAA 4176
              |||||||||||||||||||||||||
Db     128956 GTTTTAAAAATGGAGAAAGATTTAA 128932


RESULT 5
AC090055/c
LOCUS       AC090055               143959 bp    DNA     linear   HTG 13-FEB-2001
DEFINITION  Oryza sativa chromosome 9 clone PAC0651G05, *** SEQUENCING IN
            PROGRESS ***, 9 ordered pieces.
ACCESSION   AC090055
VERSION     AC090055.1  GI:12751517
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 143959)
  AUTHORS   Vanavichit,A.
  TITLE     Oryza sativa PAC P0651G05 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 143959)
  AUTHORS   Vanavichit,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-FEB-2001) Rice Genome Sequencing Project, National
            Center for Genetic Engineering and Biotechnology (BIOTEC), at
            Kasetsart University, Khamphaengsaen, Nakornpathom 73140, Thailand
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1     6613: contig of 6613 bp in length
            *     6614     6713: gap of unknown length
            *     6714    14584: contig of 7871 bp in length
            *    14585    14684: gap of unknown length
            *    14685    36811: contig of 22127 bp in length
            *    36812    36911: gap of unknown length
            *    36912    44742: contig of 7831 bp in length
            *    44743    44842: gap of unknown length
            *    44843    54340: contig of 9498 bp in length
            *    54341    54440: gap of unknown length
            *    54441    57040: contig of 2600 bp in length
            *    57041    57140: gap of unknown length
            *    57141    73511: contig of 16371 bp in length
            *    73512    73611: gap of unknown length
            *    73612    110910: contig of 37299 bp in length
            *   110911    111010: gap of unknown length
            *   111011    143959: contig of 32949 bp in length.
FEATURES             Location/Qualifiers
     source          1. .143959
                     /organism="Oryza sativa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4530"
                     /chromosome="9"
                     /clone="PAC0651G05"
BASE COUNT    42436 a   29400 c   28888 g   42434 t    801 others
ORIGIN

  Query Match             0.3%;  Score 24;  DB 2;  Length 143959;
  Best Local Similarity   100.0%;  Pred. No. 1.6;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4485 TAACTTGTACAAAATCCTTATAAA 4508
              ||||||||||||||||||||||||
Db     133945 TAACTTGTACAAAATCCTTATAAA 133922


RESULT 6
AC144923
LOCUS       AC144923               90079 bp    DNA     linear   HTG 26-MAY-2003
DEFINITION  Mus musculus chromosome 7 clone RP23-366A16 map 7, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC144923
VERSION     AC144923.1  GI:31075322
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 90079)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 7, clone RP23-366A16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 90079)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L28922
                Center clone name: 366_A_16
```

```
--------------.
* NOTE: This record contains 87 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      943: contig of 943 bp in length
*      944     1043: gap of 100 bp
*     1044     1975: contig of 932 bp in length
*     1976     2075: gap of 100 bp
*     2076     2969: contig of 894 bp in length
*     2970     3069: gap of 100 bp
*     3070     3985: contig of 916 bp in length
*     3986     4085: gap of 100 bp
*     4086     5014: contig of 929 bp in length
*     5015     5114: gap of 100 bp
*     5115     6055: contig of 941 bp in length
*     6056     6155: gap of 100 bp
*     6156     7105: contig of 950 bp in length
*     7106     7205: gap of 100 bp
*     7206     8144: contig of 939 bp in length
*     8145     8244: gap of 100 bp
*     8245     9184: contig of 940 bp in length
*     9185     9284: gap of 100 bp
*     9285    10209: contig of 925 bp in length
*    10210    10309: gap of 100 bp
*    10310    11263: contig of 954 bp in length
*    11264    11363: gap of 100 bp
*    11364    12312: contig of 949 bp in length
*    12313    12412: gap of 100 bp
*    12413    13378: contig of 966 bp in length
*    13379    13478: gap of 100 bp
*    13479    14418: contig of 940 bp in length
*    14419    14518: gap of 100 bp
*    14519    15451: contig of 933 bp in length
*    15452    15551: gap of 100 bp
*    15552    16457: contig of 906 bp in length
*    16458    16557: gap of 100 bp
*    16558    17498: contig of 941 bp in length
*    17499    17598: gap of 100 bp
*    17599    18562: contig of 964 bp in length
*    18563    18662: gap of 100 bp
*    18663    19612: contig of 950 bp in length
*    19613    19712: gap of 100 bp
*    19713    20649: contig of 937 bp in length
*    20650    20749: gap of 100 bp
*    20750    21665: contig of 916 bp in length
*    21666    21765: gap of 100 bp
*    21766    22652: contig of 887 bp in length
*    22653    22752: gap of 100 bp
*    22753    23708: contig of 956 bp in length
*    23709    23808: gap of 100 bp
*    23809    24737: contig of 929 bp in length
*    24738    24837: gap of 100 bp
*    24838    25774: contig of 937 bp in length
*    25775    25874: gap of 100 bp
*    25875    26841: contig of 967 bp in length
*    26842    26941: gap of 100 bp
*    26942    27886: contig of 945 bp in length
*    27887    27986: gap of 100 bp
*    27987    28907: contig of 921 bp in length
*    28908    29007: gap of 100 bp
*    29008    29967: contig of 960 bp in length
*    29968    30067: gap of 100 bp
*    30068    31027: contig of 960 bp in length
*    31028    31127: gap of 100 bp
*    31128    32048: contig of 921 bp in length
*    32049    32148: gap of 100 bp
*    32149    33115: contig of 967 bp in length
*    33116    33215: gap of 100 bp
*    33216    34174: contig of 959 bp in length
*    34175    34274: gap of 100 bp
*    34275    35228: contig of 954 bp in length
*    35229    35328: gap of 100 bp
*    35329    36281: contig of 953 bp in length
*    36282    36381: gap of 100 bp
*    36382    37308: contig of 927 bp in length
*    37309    37408: gap of 100 bp
*    37409    38354: contig of 946 bp in length
*    38355    38454: gap of 100 bp
*    38455    39379: contig of 925 bp in length
*    39380    39479: gap of 100 bp
*    39480    40408: contig of 929 bp in length
*    40409    40508: gap of 100 bp
*    40509    41435: contig of 927 bp in length
*    41436    41535: gap of 100 bp
*    41536    42475: contig of 940 bp in length
*    42476    42575: gap of 100 bp
*    42576    43510: contig of 935 bp in length
*    43511    43610: gap of 100 bp
*    43611    44551: contig of 941 bp in length
*    44552    44651: gap of 100 bp
*    44652    45626: contig of 975 bp in length
*    45627    45726: gap of 100 bp
*    45727    46666: contig of 940 bp in length
*    46667    46766: gap of 100 bp
*    46767    47697: contig of 931 bp in length
*    47698    47797: gap of 100 bp
*    47798    48742: contig of 945 bp in length
*    48743    48842: gap of 100 bp
*    48843    49763: contig of 921 bp in length
*    49764    49863: gap of 100 bp
*    49864    50759: contig of 896 bp in length
*    50760    50859: gap of 100 bp
*    50860    51828: contig of 969 bp in length
*    51829    51928: gap of 100 bp
*    51929    52888: contig of 960 bp in length
*    52889    52988: gap of 100 bp
*    52989    53926: contig of 938 bp in length
*    53927    54026: gap of 100 bp
*    54027    54987: contig of 961 bp in length
*    54988    55087: gap of 100 bp
*    55088    56001: contig of 914 bp in length
*    56002    56101: gap of 100 bp
*    56102    57011: contig of 910 bp in length
*    57012    57111: gap of 100 bp
*    57112    58032: contig of 921 bp in length
*    58033    58132: gap of 100 bp
*    58133    59075: contig of 943 bp in length
*    59076    59175: gap of 100 bp
*    59176    60143: contig of 968 bp in length
*    60144    60243: gap of 100 bp
*    60244    61133: contig of 890 bp in length
*    61134    61233: gap of 100 bp
*    61234    62158: contig of 925 bp in length
*    62159    62258: gap of 100 bp
*    62259    63197: contig of 939 bp in length
*    63198    63297: gap of 100 bp
*    63298    64244: contig of 947 bp in length
*    64245    64344: gap of 100 bp
*    64345    65294: contig of 950 bp in length
*    65295    65394: gap of 100 bp
*    65395    66354: contig of 960 bp in length
*    66355    66454: gap of 100 bp
*    66455    67419: contig of 965 bp in length
*    67420    67519: gap of 100 bp
*    67520    68447: contig of 928 bp in length
*    68448    68547: gap of 100 bp
*    68548    69455: contig of 908 bp in length
*    69456    69555: gap of 100 bp
```

```
          *    69556    70463: contig of 908 bp in length
          *    70464    70563: gap of 100 bp
          *    70564    71465: contig of 902 bp in length
          *    71466    71565: gap of 100 bp
          *    71566    72508: contig of 943 bp in length

  Query Match             0.3%;  Score 23;  DB 2;  Length 90079;
  Best Local Similarity   100.0%;  Pred. No. 5.9;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6010 TCTGAATCTATAGTTACATCTTT 6032
             |||||||||||||||||||||||
Db      37554 TCTGAATCTATAGTTACATCTTT 37576


RESULT 7
AC124702
LOCUS       AC124702              105183 bp    DNA     linear   ROD 16-NOV-2002
DEFINITION  Mus musculus chromosome 7 clone RP24-254J9, complete sequence.
ACCESSION   AC124702
VERSION     AC124702.3  GI:25046669
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 105183)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 105183)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUN-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 105183)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 105183)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Nov 16, 2002 this sequence version replaced gi:24371518.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information -----------------
            Center project name: M_BB0254J09
            ------------------------------------------------------
FEATURES             Location/Qualifiers
     source          1. .105183
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="7"
                     /clone="RP24-254J9"
BASE COUNT    32179 a   21512 c   20581 g   30911 t
ORIGIN

  Query Match             0.3%;  Score 23;  DB 10;  Length 105183;
  Best Local Similarity   100.0%;  Pred. No. 6;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6010 TCTGAATCTATAGTTACATCTTT 6032
             |||||||||||||||||||||||
Db      23074 TCTGAATCTATAGTTACATCTTT 23096


RESULT 8
AP003394
LOCUS       AP003394              164778 bp    DNA     linear   PRI 15-MAR-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11q clone:RP11-278J24,
            complete sequences.
ACCESSION   AP003394
VERSION     AP003394.1  GI:20302476
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2002)
REFERENCE   2  (bases 1 to 164778)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2001) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     This work was done in  collaboration with Arai, Y., Kubo, T. and
            Ohki, M.
            National Cancer Center Research Institute
            Cancer Genomic Division
            5-1-1, Tsukiji, Chuo-ku, Tokyo, JAPAN
            zip: 104-0045
            phone: 81-3-3542-2511 ex 4752, fax: 81-3-3542-0688 e-mail:
            yarai@ncc.go.jp.
FEATURES             Location/Qualifiers
     source          1. .164778
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-278J24"
BASE COUNT    53727 a   27854 c   29413 g   53784 t
ORIGIN

  Query Match             0.3%;  Score 23;  DB 9;  Length 164778;
  Best Local Similarity   100.0%;  Pred. No. 6.1;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1356 TTTAAATGGAAAAACAAAAAAAT 1378
             |||||||||||||||||||||||
Db     151487 TTTAAATGGAAAAACAAAAAAAT 151509


RESULT 9
AC131111
LOCUS       AC131111              179068 bp    DNA     linear   ROD 17-MAY-2003
DEFINITION  Mus musculus chromosome 7 clone RP23-449N4, complete sequence.
ACCESSION   AC131111
VERSION     AC131111.4  GI:30841068
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 179068)
  AUTHORS   Wilson,R.K.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179068)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (16-AUG-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 179068)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-APR-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 179068)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On May 17, 2003 this sequence version replaced gi:29788934.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information ----------------
            Center project name: M_BA0449N04
            -----------------------------------------------------
FEATURES             Location/Qualifiers
     source          1. .179068
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="7"
                     /clone="RP23-449N4"
BASE COUNT    54368 a  36773 c  34978 g  52949 t
ORIGIN

  Query Match             0.3%;  Score 23;  DB 10;  Length 179068;
  Best Local Similarity   100.0%;  Pred. No. 6.1;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6010 TCTGAATCTATAGTTACATCTTT 6032
              |||||||||||||||||||||||
Db      20992 TCTGAATCTATAGTTACATCTTT 21014


RESULT 10
AP003558
LOCUS       AP003558               182975 bp    DNA     linear   PRI 15-MAR-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11q clone:RP11-693A6, complete
            sequences.
ACCESSION   AP003558
VERSION     AP003558.2  GI:17939971
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 182975)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2001) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Dec 19, 2001 this sequence version replaced gi:13928407.
FEATURES             Location/Qualifiers
     source          1. .182975
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-693A6"
BASE COUNT    58852 a  31033 c  33120 g  59970 t
ORIGIN

  Query Match             0.3%;  Score 23;  DB 9;  Length 182975;
  Best Local Similarity   100.0%;  Pred. No. 6.1;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1356 TTTAAATGGAAAAACAAAAAAAT 1378
              |||||||||||||||||||||||
Db       4905 TTTAAATGGAAAAACAAAAAAAT 4927


RESULT 11
AC027247
LOCUS       AC027247               188596 bp    DNA     linear   HTG 27-APR-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-743I15 map 11, WORKING DRAFT
            SEQUENCE, 21 unordered pieces.
ACCESSION   AC027247
VERSION     AC027247.2  GI:7651926
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 188596)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-743I15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 188596)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 27, 2000 this sequence version replaced gi:7331617.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L8857
                Center clone name: 743_I_15
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 177981 bases at least Q40
```

```
            Consensus quality: 182986 bases at least Q30
            Consensus quality: 185018 bases at least Q20
            Insert size: 190000; agarose-fp
            Insert size: 186596; sum-of-contigs
            Quality coverage: 4.2 in Q20 bases; agarose-fp
            Quality coverage: 4.3 in Q20 bases; sum-of-contigs
          ------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 21 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1      228: contig of 228 bp in length
          *      229      328: gap of 100 bp
          *      329     1700: contig of 1372 bp in length
          *     1701     1800: gap of 100 bp
          *     1801     3656: contig of 1856 bp in length
          *     3657     3756: gap of 100 bp
          *     3757     5517: contig of 1761 bp in length
          *     5518     5617: gap of 100 bp
          *     5618     7498: contig of 1881 bp in length
          *     7499     7598: gap of 100 bp
          *     7599    10369: contig of 2771 bp in length
          *    10370    10469: gap of 100 bp
          *    10470    14471: contig of 4002 bp in length
          *    14472    14571: gap of 100 bp
          *    14572    18806: contig of 4235 bp in length
          *    18807    18906: gap of 100 bp
          *    18907    25234: contig of 6328 bp in length
          *    25235    25334: gap of 100 bp
          *    25335    31619: contig of 6285 bp in length
          *    31620    31719: gap of 100 bp
          *    31720    37354: contig of 5635 bp in length
          *    37355    37454: gap of 100 bp
          *    37455    46150: contig of 8696 bp in length
          *    46151    46250: gap of 100 bp
          *    46251    53881: contig of 7631 bp in length
          *    53882    53981: gap of 100 bp
          *    53982    64200: contig of 10219 bp in length
          *    64201    64300: gap of 100 bp
          *    64301    72473: contig of 8173 bp in length
          *    72474    72573: gap of 100 bp
          *    72574    84058: contig of 11485 bp in length
          *    84059    84158: gap of 100 bp
          *    84159   101127: contig of 16969 bp in length
          *   101128   101227: gap of 100 bp
          *   101228   118622: contig of 17395 bp in length
          *   118623   118722: gap of 100 bp
          *   118723   134484: contig of 15762 bp in length
          *   134485   134584: gap of 100 bp
          *   134585   153315: contig of 18731 bp in length
          *   153316   153415: gap of 100 bp
          *   153416   188596: contig of 35181 bp in length.
FEATURES             Location/Qualifiers
     source          1. .188596
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-743I15"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .228
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    329. .1700
                     /note="assembly_fragment"
     misc_feature    1801. .3656
                     /note="assembly_fragment"
     misc_feature    3757. .5517
                     /note="assembly_fragment"
     misc_feature    5618. .7498
                     /note="assembly_fragment"
     misc_feature    7599. .10369
                     /note="assembly_fragment"
     misc_feature    10470. .14471
                     /note="assembly_fragment"
     misc_feature    14572. .18806
                     /note="assembly_fragment"
     misc_feature    18907. .25234
                     /note="assembly_fragment"
     misc_feature    25335. .31619
                     /note="assembly_fragment"
     misc_feature    31720. .37354
                     /note="assembly_fragment"
     misc_feature    37455. .46150
                     /note="assembly_fragment"
     misc_feature    46251. .53881
                     /note="assembly_fragment"
     misc_feature    53982. .64200
                     /note="assembly_fragment"
     misc_feature    64301. .72473
                     /note="assembly_fragment"
     misc_feature    72574. .84058
                     /note="assembly_fragment"
     misc_feature    84159. .101127
                     /note="assembly_fragment"
     misc_feature    101228. .118622
                     /note="assembly_fragment"
     misc_feature    118723. .134484
                     /note="assembly_fragment"
     misc_feature    134585. .153315
                     /note="assembly_fragment"
     misc_feature    153416. .188596
                     /note="assembly_fragment"
BASE COUNT    59769 a  32287 c  32539 g  61997 t   2004 others
ORIGIN

  Query Match             0.3%;  Score 23;  DB 2;  Length 188596;
  Best Local Similarity   100.0%;  Pred. No. 6.1;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1356 TTTAAATGGAAAAACAAAAAAAT 1378
              |||||||||||||||||||||||
Db      99620 TTTAAATGGAAAAACAAAAAAAT 99642


RESULT 12
AC084833
LOCUS       AC084833              188922 bp    DNA     linear   HTG 22-NOV-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-1084M13 map 11, WORKING DRAFT
            SEQUENCE, 11 unordered pieces.
ACCESSION   AC084833
VERSION     AC084833.1  GI:11276173
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 188922)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-1084M13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 188922)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
```

```
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L10846
                Center clone name: 1084_M_13
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 181297 bases at least Q40
                Consensus quality: 185666 bases at least Q30
                Consensus quality: 187052 bases at least Q20
                Insert size: 190000; agarose-fp
                Insert size: 187922; sum-of-contigs
                Quality coverage: 5.6 in Q20 bases; agarose-fp
                Quality coverage: 5.6 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    10662: contig of 10662 bp in length
            *    10663    10762: gap of 100 bp
            *    10763    12511: contig of 1749 bp in length
            *    12512    12611: gap of 100 bp
            *    12612    16569: contig of 3958 bp in length
            *    16570    16669: gap of 100 bp
            *    16670    21154: contig of 4485 bp in length
            *    21155    21254: gap of 100 bp
            *    21255    71078: contig of 49824 bp in length
            *    71079    71178: gap of 100 bp
            *    71179    86206: contig of 15028 bp in length
            *    86207    86306: gap of 100 bp
            *    86307   107726: contig of 21420 bp in length
            *   107727   107826: gap of 100 bp
            *   107827   125885: contig of 18059 bp in length
            *   125886   125985: gap of 100 bp
            *   125986   150510: contig of 24525 bp in length
            *   150511   150610: gap of 100 bp
            *   150611   182811: contig of 32201 bp in length
            *   182812   182911: gap of 100 bp
            *   182912   188922: contig of 6011 bp in length.
FEATURES            Location/Qualifiers
     source         1. .188922
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="11"
                    /map="11"
                    /clone="RP11-1084M13"
                    /clone_lib="RPCI-11 Human Male BAC"
     misc_feature   1. .10662
                    /note="assembly_fragment
                    clone_end:SP6
                    vector_side:left"
     misc_feature   10763. .12511
                    /note="assembly_fragment"
     misc_feature   12612. .16569
                    /note="assembly_fragment"
     misc_feature   16670. .21154
                    /note="assembly_fragment"
     misc_feature   21255. .71078
                    /note="assembly_fragment"
     misc_feature   71179. .86206
                    /note="assembly_fragment"
     misc_feature   86307. .107726
                    /note="assembly_fragment"
     misc_feature   107827. .125885
                    /note="assembly_fragment"
     misc_feature   125986. .150510
                    /note="assembly_fragment"
     misc_feature   150611. .182811
                    /note="assembly_fragment"
     misc_feature   182912. .188922
                    /note="assembly_fragment
                    clone_end:T7
                    vector_side:right"
BASE COUNT    61951 a  32734 c  33021 g  60207 t   1009 others
ORIGIN

  Query Match             0.3%;  Score 23;  DB 2;  Length 188922;
  Best Local Similarity   100.0%;  Pred. No. 6.1;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1356 TTTAAATGGAAAAACAAAAAAAT 1378
              |||||||||||||||||||||||
Db       4908 TTTAAATGGAAAAACAAAAAAAT 4930


RESULT 13
AP004249
LOCUS       AP004249              191109 bp    DNA     linear   PRI 24-JAN-2002
DEFINITION  Homo sapiens genomic DNA, chromosome 11q clone:RP11-743I15,
            complete sequences.
ACCESSION   AP004249
VERSION     AP004249.2  GI:18307740
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 191109)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-2001) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Jan 23, 2002 this sequence version replaced gi:16117846.
FEATURES            Location/Qualifiers
     source         1. .191109
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
```

```
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-743I15"
BASE COUNT    61710 a  32427 c  34112 g  62860 t
ORIGIN

  Query Match             0.3%;  Score 23;  DB 9;  Length 191109;
  Best Local Similarity   100.0%;  Pred. No. 6.1;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1356 TTTAAATGGAAAAACAAAAAAAT 1378
              |||||||||||||||||||||||
Db      99895 TTTAAATGGAAAAACAAAAAAAT 99917


RESULT 14
AC128738
LOCUS       AC128738               195779 bp    DNA     linear   HTG 25-AUG-2002
DEFINITION  Mus musculus chromosome UNK clone RP24-106M4, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC128738
VERSION     AC128738.2  GI:22475847
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 195779)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 195779)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 195779)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Aug 25, 2002 this sequence version replaced gi:21930275.

            ----------------- Genome Center ----------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information -------------------
            Center project name: M_BB0106M04
            ------------------------------------------------------
            -------------- Summary Statistics  -------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 192403 bases at least Q40
            Consensus quality: 193021 bases at least Q30
            Consensus quality: 193330 bases at least Q20
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1550: contig of 1550 bp in length
            *     1551     1650: gap of unknown length
            *     1651     7955: contig of 6305 bp in length
            *     7956     8055: gap of unknown length
            *     8056    17849: contig of 9794 bp in length
            *    17850    17949: gap of unknown length
            *    17950    34634: contig of 16685 bp in length
            *    34635    34734: gap of unknown length
            *    34735    56089: contig of 21355 bp in length
            *    56090    56189: gap of unknown length
            *    56190   128179: contig of 71990 bp in length
            *   128180   128279: gap of unknown length
            *   128280   194301: contig of 66022 bp in length
            *   194302   194401: gap of unknown length
            *   194402   194498: contig of 97 bp in length
            *   194499   194598: gap of unknown length
            *   194599   195779: contig of 1181 bp in length.
FEATURES             Location/Qualifiers
     source          1. .195779
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP24-106M4"
     misc_feature    1. .1550
                     /note="assembly_name:Contig10"
     misc_feature    1651. .7955
                     /note="assembly_name:Contig11"
     misc_feature    8056. .17849
                     /note="assembly_name:Contig12"
     misc_feature    17950. .34634
                     /note="assembly_name:Contig13"
     misc_feature    34735. .56089
                     /note="assembly_name:Contig14"
     misc_feature    56190. .128179
                     /note="assembly_name:Contig15"
     misc_feature    128280. .194301
                     /note="assembly_name:Contig16"
     misc_feature    194402. .194498
                     /note="assembly_name:Contig5"
     misc_feature    194599. .195779
                     /note="assembly_name:Contig9"
BASE COUNT    57716 a  40158 c  38829 g  58264 t    812 others
ORIGIN

  Query Match             0.3%;  Score 23;  DB 2;  Length 195779;
  Best Local Similarity   100.0%;  Pred. No. 6.1;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6010 TCTGAATCTATAGTTACATCTTT 6032
              |||||||||||||||||||||||
Db     148591 TCTGAATCTATAGTTACATCTTT 148613


RESULT 15
AC128738/c
LOCUS       AC128738               195779 bp    DNA     linear   HTG 25-AUG-2002
DEFINITION  Mus musculus chromosome UNK clone RP24-106M4, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC128738
VERSION     AC128738.2  GI:22475847
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 195779)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 195779)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 195779)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Aug 25, 2002 this sequence version replaced gi:21930275.

            ------------------- Genome Center ---------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information -------------------
            Center project name: M_BB0106M04
            ------------------------------------------------------
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 192403 bases at least Q40
            Consensus quality: 193021 bases at least Q30
            Consensus quality: 193330 bases at least Q20
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1550: contig of 1550 bp in length
            *     1551     1650: gap of unknown length
            *     1651     7955: contig of 6305 bp in length
            *     7956     8055: gap of unknown length
            *     8056    17849: contig of 9794 bp in length
            *    17850    17949: gap of unknown length
            *    17950    34634: contig of 16685 bp in length
            *    34635    34734: gap of unknown length
            *    34735    56089: contig of 21355 bp in length
            *    56090    56189: gap of unknown length
            *    56190   128179: contig of 71990 bp in length
            *   128180   128279: gap of unknown length
            *   128280   194301: contig of 66022 bp in length
            *   194302   194401: gap of unknown length
            *   194402   194498: contig of 97 bp in length
            *   194499   194598: gap of unknown length
            *   194599   195779: contig of 1181 bp in length.
FEATURES             Location/Qualifiers
     source          1. .195779
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP24-106M4"
     misc_feature    1. .1550
                     /note="assembly_name:Contig10"
     misc_feature    1651. .7955
                     /note="assembly_name:Contig11"
     misc_feature    8056. .17849
                     /note="assembly_name:Contig12"
     misc_feature    17950. .34634
                     /note="assembly_name:Contig13"
     misc_feature    34735. .56089
                     /note="assembly_name:Contig14"
     misc_feature    56190. .128179
                     /note="assembly_name:Contig15"
     misc_feature    128280. .194301
                     /note="assembly_name:Contig16"
     misc_feature    194402. .194498
                     /note="assembly_name:Contig5"
     misc_feature    194599. .195779
                     /note="assembly_name:Contig9"

BASE COUNT    57716 a  40158 c  38829 g  58264 t    812 others
ORIGIN

  Query Match             0.3%;  Score 23;  DB 2;  Length 195779;
  Best Local Similarity   100.0%;  Pred. No. 6.1;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6010 TCTGAATCTATAGTTACATCTTT 6032
              |||||||||||||||||||||||
Db      16923 TCTGAATCTATAGTTACATCTTT 16901


RESULT 16
ATCHRIV46
LOCUS       ATCHRIV46             198788 bp    DNA     linear   PLN 16-MAR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 4, contig fragment No. 46.
ACCESSION   AL161546
VERSION     AL161546.2  GI:7268471
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 198788)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATCHRIV45 at the 5' end and an
            overlap with ATCHRIV47 at the 3' end.
FEATURES             Location/Qualifiers
     source          1. .198788
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            5243. .6388
                     /gene="AT4g17200"
     CDS             5243. .6388
                     /gene="AT4g17200"
                     /note="similarity to several hypothetical proteins
                     contains EST gb:AI995149.1, H76318, AA651584, T42945,
                     T22973, AI997891.1, AI995403.1"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB78723.1"
                     /db_xref="GI:7268472"
                     /translation="MTTMSDLSPDLVGEILTRVPMTSLISVRCTCKMWNALSKEGIFF
                     KAARKQFMGFTMMDSRVCSMKFDLQGIRNNEHDFVDPCIKQIAKLDQIEVSKVLQCDG
                     LLLCVGKDNSRLLVWNPYLGQTRFIKPRKRFNKLEWYALGYDNNHNYKILMNYDTGHL
                     FGYEIYDFSSDSWRVLDVSPDCHIPFHQQSVSLNGNTYFLAQEKIIVEGEEEEVVDEI
                     EKFLLCFDFTTERFGPRLPLPFHSDVLETVAISCVRDDQLAVLYQRFETWEIWVTTKI
                     DPTAVSWSMFLSVDMEPPTGCQFDDEAGSFFIDEENKVAVFFNSDLFNPRVENRNRRY
                     PTAYIIGQDGYFKSATLREAPDLVKPDPYVYCCPLVWSSYAPSLVQL"
     exon            5243. .6388
                     /gene="AT4g17200"
                     /number=1
     gene            6720. .8392
                     /gene="AT4g17210"
     gene            complement(join(6720. .8114,8204. .8392))
                     /gene="AT4g17210"
     CDS             complement(join(6720. .8114,8204. .8392))
                     /gene="AT4g17210"
                     /note="similarity to myosin heavy chain (contains myosin
```

ATPase (EC 3.6.1.32)), skeletal muscle, embryonic, rat"
/codon_start=1
/product="putative protein"
/protein_id="CAB78724.1"
/db_xref="GI:7268473"
/translation="MAKIRTDAPVMPPETPPRSSEVGEIDTRAPFQSVRDALSLFRQV
SFSKKQPPRLSSSSSSQSQDTTTDVSDKEMQLLLAEQEMDRVKICLDGSVAAKAQALS
DLDSAQRKAADLRVKLESIKHSRKCAISTKHTMNQRLEQLQSENQETESTREDYILIT
AELFMAKYELAELKQQFNLSVEERLAELQRAEEAECASMVNSNKIKDMSHDIAEMRDA
AERLNSDAARKKEEEEQIKEESIALRETYVCKKLEAKQRLEDLKRDCDPELKKDIEEL
MEISTENERLQEEIKLSGELKEAKSAMQEIYDEESSYKSLVGSLTVELDGVQRENREL
KGKEKERQEAEEGEWVEASRKVDEIMREAEKTRKEAEEMRMNVDELRREAAAKHMVMG
EAVKQLEIVGRAVEKAKTAEKRAVEDMKVLTEKKESLTHDEPDKKIRISLKEYEELRG
KHEESERMVQFKAKTVAAQLEEINESRIEGERKLEEKIKEMEELKAAIDGALRKAEIA
EEAHSIVDAELRKWKPQDL"
exon complement(6720. .8114)
/gene="AT4g17210"
/number=1
intron complement(8115. .8203)
/gene="AT4g17210"
/number=1
exon complement(8204. .8392)
/gene="AT4g17210"
/number=2
repeat_region 9263. .9287
gene 11143. .13540
/gene="AT4g17220"
gene complement(join(11143. .11170,11275. .11414,11502. .11933,
11990. .12073,12166. .12312,12389. .12721,12806. .12976,
13076. .13141,13298. .13383,13480. .13540))
/gene="AT4g17220"
CDS complement(join(11143. .11170,11275. .11414,11502. .11933,
11990. .12073,12166. .12312,12389. .12721,12806. .12976,
13076. .13141,13298. .13383,13480. .13540))
/gene="AT4g17220"
/note="similarity to myosin beta heavy chain, cardiac and
skeletal muscle - Homo sapiens, PIR:S12458
contains EST gb:AI998577.1"
/codon_start=1
/product="putative protein"
/protein_id="CAB78725.1"
/db_xref="GI:7268474"
/translation="MTAAENPFVSDTSSLQSQLKEKEKELLAAKAEVEALRTNEELKD
RVFKELRENVRKLEEKLGATENQVDQKELERKKLEEEKEDALAAQDAAEEALRRVYTH
QQDDDSLPLESIIAPLESQIKIHKHEISALQEDKKALERLTKSKESALLEAERILRSA
LERALIVEEVQNHNFELRRQIEICQDENKFLEKINRQKVLEIEKLSQSIVELEEAILA
GGTAANAVRDYRRQISQLNASDEKRTLERELARVKVSASRVALAVANEWKDENDRVMP
VKQWLEERRILHGEMQKLKDKLAVSERTAKAESQLKVIAYERLKLRLKTIEDGLKGPN
TFFVSPTTKTEKSGKILGFLTSGGGSKKRSSSQLRGSVTGRIHAMNQPIDRVGESDEM
ENSKITANGLTDQHEEDSERKTEEDGNVYSEDMVSGFLYDRLQKEVIALRKLCESKEG
TINAKNEEIKKVDALTKAIEVETKKAKREAAARE KENALAMLNEESKLCRKAKLPRSR
IPNPRCQ"
exon complement(11143. .11170)
/gene="AT4g17220"
/number=1
intron complement(11171. .11274)
/gene="AT4g17220"
/number=1
exon complement(11275. .11414)
/gene="AT4g17220"
/number=2
intron complement(11415. .11501)
/gene="AT4g17220"
/number=2
exon complement(11502. .11933)
/gene="AT4g17220"
/number=3
intron complement(11934. .11989)
/gene="AT4g17220"
/number=3
exon complement(11990. .12073)
/gene="AT4g17220"
/number=4
intron complement(12074. .12165)
/gene="AT4g17220"
/number=4
exon complement(12166. .12312)
/gene="AT4g17220"
/number=5
intron complement(12313. .12388)
/gene="AT4g17220"
/number=5
exon complement(12389. .12721)
/gene="AT4g17220"
/number=6
intron complement(12722. .12805)
/gene="AT4g17220"
/number=6
exon complement(12806. .12976)
/gene="AT4g17220"
/number=7
intron complement(12977. .13075)
/gene="AT4g17220"
/number=7
exon complement(13076. .13141)
/gene="AT4g17220"
/number=8
intron complement(13142. .13297)
/gene="AT4g17220"
/number=8
exon complement(13298. .13383)
/gene="AT4g17220"
/number=9
intron complement(13384. .13479)
/gene="AT4g17220"
/number=9
exon complement(13480. .13540)
/gene="AT4g17220"
/number=10
gene 15529. .16392
/gene="AT4g17230"
gene complement(15529. .16392)
/gene="AT4g17230"
CDS complement(15529. .16392)
/gene="AT4g17230"
/note="contains EST gb:N37425, AA720344, R29917"
/codon_start=1
/product="scarecrow-like 13 (SCL13)"
/protein_id="CAB78726.1"
/db_xref="GI:7268475"
/translation="MVSVSGSPIQRLGTYMAEGLRARLEGSGSNIYKSLKCNEPTGRE
LMSYMSVLYEICPYWKFAYTTANVEILEAIAGETRVHIIDFQIAQGSQYMFLIQELAK
HPGGPPLLRVTGVDDSQSTYARGGGLSLVGERLATLAQSCGVPFEFHDAIMSGCKVQR
EHLGLEPGFAVVVNFPYVLHHMPDESVSVENHRDRLLHLIKSLSPKLVTLVEQESNTN
TSPFLSRFVETLDYYTAMFESIDAARPRDDKQRISAEQHCVARDIVNMIACEESERVE
RHEVLGNGGSE"
exon complement(15529. .16392)
/gene="AT4g17230"
/number=1
gene 20934. .23722
/gene="AT4g17240"
gene complement(join(20934. .21113,21209. .21292,21390. .21458,
21610. .21813,21933. .21983,22061. .22306,23600. .23722))
/gene="AT4g17240"
CDS complement(join(20934. .21113,21209. .21292,21390. .21458,
21610. .21813,21933. .21983,22061. .22306,23600. .23722))
/gene="AT4g17240"
/note="contains EST gb:Z48426"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB78727.1"

Query Match 0.3%; Score 23; DB 8; Length 198788;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1354 TCTTTAAATGGAAAAACAAAAAA 1376
              ||||||||||||||||||||||
Db     187402 TCTTTAAATGGAAAAACAAAAAA 187424


RESULT 17
ATFCA8
LOCUS       ATFCA8                207674 bp    DNA     linear   PLN 27-AUG-1999
DEFINITION  Arabidopsis thaliana DNA chromosome 4, ESSA I FCA contig fragment
            No. 8.
ACCESSION   Z97343
VERSION     Z97343.1  GI:2245073
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Bevan,M., Stiekema,W., Murphy,G., Wambutt,R., Pohl,T., Terryn,N.,
            Kreis,M., Kavanagh,T., Entian,K.D., Rieger,M., James,R.,
            Puigdomenech,P., Hatzopoulos,P., Obermaier,B., Duesterhoft,A.,
            Jones,J., Palme,K., Ansorge,W., Delseny,M., Bancroft,I.,
            Mewes,H.W., Schueller,C. and Chalwatzis,N.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 207674)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1999) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            schuelle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATFCA7 at the 5' end and an
            overlap with ATFCA9 at the 3' end.
FEATURES             Location/Qualifiers
     source          1. .207674
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     source          1. .19830
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="cosmid G11319"
     source          11329. .50912
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="cosmid G19830"
     source          49076. .58338
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="cosmid G2484"
     source          57182. .66642
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="cosmid G7006"
     source          66642. .89059
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="cosmid CC29F3"
     source          89059. .116201
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="cosmid CC18E16"
     source          116201. .142160
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="cosmid CC9C4/CC22C20"
     source          142160. .148135
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="cosmid YUP24F4.IB12F15"
     source          148135. .192614
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="cosmid YUP24F4.IB13H3/G3845"
     source          190619. .207674
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="BAC TAMU3P6"
     gene            3388. .6418
                     /gene="dl4610w"
     CDS             join(3388. .3428,3488. .3594,3913. .4025,4104. .4198,
                     4932. .5298,5556. .5693,5989. .6117,6200. .6418)
                     /gene="dl4610w"
                     /note="weak similarity to 319K protein ndvB, Rhizobium
                     meliloti
                     contains EST gb:947762"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB10496.1"
                     /db_xref="GI:2245074"
                     /db_xref="SPTREMBL:O23560"
                     /translation="MIDQFINFVIRPPSCWGCFDRAEYDPDQYLWEKEFSLGGTKCKR
                     QDLEANANEAVMVLLPSNITVFTLDFSGSGLSEGDYVSLGWHEKDDLKTVVSYLRNSN
                     QVSRIGLWGRSMGAVTRFRPRLLFQLYLGMQAETNSSNLITLTSFSSAMRYLHNCGFY
                     VQAYQSFFIFMYLSSGNLQGDKNIIKFDGDHNSSRPQSYYDSVLVFFYNVLRPPPISS
                     SYSSKLESYYSLGDVNSATGLDESFLYEIISGLRSACIDVASSSSAPPAPLTTKPTNE
                     LLSEAMPMIDTGKRIDQFEEGCSFTSSNRESWGRCSSLGGTEEDESLTAGEGDQVEKT
                     ADVNTERKPRDSSREEEEEDSKEKKIKNGGETDAKKPRHEKLERLEAFSKRLRHYILK
                     RVNHRRHRSP"
     exon            3388. .3428
                     /gene="dl4610w"
                     /number=1
     intron          3429. .3487
                     /gene="dl4610w"
                     /number=1
     exon            3488. .3594
                     /gene="dl4610w"
                     /number=2
     intron          3595. .3912
                     /gene="dl4610w"
                     /number=2
     exon            3913. .4025
                     /gene="dl4610w"
                     /number=3
     intron          4026. .4103
                     /gene="dl4610w"
                     /number=3
     exon            4104. .4198
                     /gene="dl4610w"
                     /number=4
     intron          4199. .4931
                     /gene="dl4610w"
                     /number=4
     exon            4932. .5298
                     /gene="dl4610w"
                     /number=5
     intron          5299. .5555
                     /gene="dl4610w"
```

```
                     /number=5
     exon            5556. .5693
                     /gene="dl4610w"
                     /number=6
     intron          5694. .5988
                     /gene="dl4610w"
                     /number=6
     exon            5989. .6117
                     /gene="dl4610w"
                     /number=7
     intron          6118. .6199
                     /gene="dl4610w"
                     /number=7
     exon            6200. .6418
                     /gene="dl4610w"
                     /number=8
     gene            6750. .8311
                     /gene="dl4615c"
     gene            complement(join(6750. .6893,7083. .7287,7618. .7700,
                     7990. .8129,8266. .8311))
                     /gene="dl4615c"
     CDS             complement(join(6750. .6893,7083. .7287,7618. .7700,
                     7990. .8129,8266. .8311))
                     /gene="dl4615c"
                     /note="strong similarity to RAB2A, Lotus japonicus"
                     /codon_start=1
                     /product="GTP-binding RAB2A like protein"
                     /protein_id="CAB10497.1"
                     /db_xref="GI:2245075"
                     /db_xref="SPTREMBL:O23561"
                     /translation="MSYAYRFKYIIIGDTGVGKSCLLLKFTDKRFQAVHDLTIGVEFG
                     AKTITIDNKPIKLQIWDTAGQESFRSVTRSYYRGRAGTLLVYDITRRETFNHLASWLE
                     EARQHASENMTTMLIGNKCDLEDKRTVSTEEGEQFAREHGLIFMEASAKTAHNVEEAF
                     VETAATIYKRIQDGVVDEANEPGITPGPFGGKDASSSQQRRGCCG"
     exon            complement(6750. .6893)
                     /gene="dl4615c"
                     /number=1
     intron          complement(6894. .7082)
                     /gene="dl4615c"
                     /number=1
     exon            complement(7083. .7287)
                     /gene="dl4615c"
                     /number=2
     intron          complement(7288. .7617)
                     /gene="dl4615c"
                     /number=2
     exon            complement(7618. .7700)
                     /gene="dl4615c"
                     /number=3
     intron          complement(7701. .7989)
                     /gene="dl4615c"
                     /number=3
     exon            complement(7990. .8129)
                     /gene="dl4615c"
                     /number=4
     intron          complement(8130. .8265)
                     /gene="dl4615c"
                     /number=4
     exon            complement(8266. .8311)
                     /gene="dl4615c"
                     /number=5

  Query Match             0.3%;  Score 23;  DB 8;  Length 207674;
  Best Local Similarity   100.0%;  Pred. No. 6.2;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1354 TCTTTAAATGGAAAAACAAAAAA 1376
             |||||||||||||||||||||||
Db     198037 TCTTTAAATGGAAAAACAAAAAA 198059


RESULT 18
AC129867
```

```
LOCUS       AC129867               227486 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-107A6, *** SEQUENCING IN PROGRESS
            ***.
ACCESSION   AC129867
VERSION     AC129867.3  GI:30578824
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 227486)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 227486)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 227486)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
```

```
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:23664630.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GJOX
                Center clone name: CH230-107A6
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 221947 bases at least Q40
                Consensus quality: 223543 bases at least Q30
                Consensus quality: 224589 bases at least Q20
                Estimated insert size: 235363; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   227486: contig of 227486 bp in length.
FEATURES             Location/Qualifiers
     source          1. .227486
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-107A6"
     misc_feature    1. .2872
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    3737. .4577
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH358208"
     misc_feature    complement(226777. .227441)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH358207"
BASE COUNT    72274 a  41711 c  40817 g  70640 t   2044 others
ORIGIN

  Query Match             0.3%;  Score 23;  DB 2;  Length 227486;
  Best Local Similarity   100.0%;  Pred. No. 6.2;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2372 AAATAAATAAAAACAAGCAAATG 2394
             |||||||||||||||||||||||
Db     106490 AAATAAATAAAAACAAGCAAATG 106512
```

```
RESULT 19
AC106122/c
LOCUS       AC106122              278353 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-115P21, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC106122
VERSION     AC106122.4  GI:30521569
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 278353)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 278353)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 278353)
```

```
  AUTHORS'  Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:23101777.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GHLE
                Center clone name: CH230-115P21
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 246842 bases at least Q40
                Consensus quality: 249314 bases at least Q30
                Consensus quality: 251061 bases at least Q20
                Estimated insert size: 255392; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   126270: contig of 126270 bp in length
            *  126271   126370: gap of unknown length
            *  126371   273996: contig of 147626 bp in length
            *  273997   274096: gap of unknown length
            *  274097   277111: contig of 3015 bp in length
            *  277112   277211: gap of unknown length
            *  277212   278353: contig of 1142 bp in length.
FEATURES             Location/Qualifiers
     source          1. .278353
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-115P21"
     misc_feature    complement(32533. .32660)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH320995"
     misc_feature    121557. .122354
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH320994"
     misc_feature    124595. .126270
                     /note="wgs_end_extension
                     clone_end:T7"
BASE COUNT    67363 a  59530 c  59718 g  65961 t  25781 others
ORIGIN
```

```
  Query Match             0.3%;  Score 23;  DB 2;  Length 278353;
  Best Local Similarity   100.0%;  Pred. No. 6.3;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5754 ACACTGCTGCTTGTGAGGATTCA 5776
              |||||||||||||||||||||||
Db     193047 ACACTGCTGCTTGTGAGGATTCA 193025


RESULT 20
MMU249416/c
LOCUS       MMU249416               1879 bp    DNA     linear   ROD 02-NOV-2000
DEFINITION  Mus musculus partial PI3Kgamma gene for phosphoinositide 3-kinase
            gamma, exon 6.
ACCESSION   AJ249416
VERSION     AJ249416.1  GI:7688233
KEYWORDS    phosphoinositide 3-kinase gamma; PI3Kgamma gene.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Hirsch,E., Wymann,M.P., Patrucco,E., Tolosano,E.,
            Bulgarelli-Leva,G., Marengo,S., Rocchi,M. and Altruda,F.
  TITLE     Analysis of the murine phosphoinositide 3-kinase gamma gene
  JOURNAL   Gene 256 (1-2), 69-81 (2000)
  MEDLINE   20509993
   PUBMED   11054537
REFERENCE   2  (bases 1 to 1879)
  AUTHORS   Hirsch,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-1999) Hirsch E., Department of Genetics, Biology
            and Biochemistry, University of Torino, Via Santena 5bis Torino,
            10126, ITALY
COMMENT     Related sequence: Joined CDS in AJ249413.
FEATURES             Location/Qualifiers
     source          1. .1879
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="12"
                     /map="band B"
     gene            1. .1879
                     /gene="PI3Kgamma"
     intron          <1. .699
                     /gene="PI3Kgamma"
                     /number=5
     exon            700. .846
                     /gene="PI3Kgamma"
                     /number=6
                     /usedin=AJ249413:P13Kg_CDS
                     /usedin=AJ249413:P13Kg_mRNA
     intron          847. .>1879
                     /gene="PI3Kgamma"
                     /number=6
BASE COUNT      515 a    395 c    377 g    588 t      4 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 10;  Length 1879;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2391 AATGGAAAACATTGTGTTTTTA 2412
              ||||||||||||||||||||||
Db       1505 AATGGAAAACATTGTGTTTTTA 1484


RESULT 21
AX346822/c
LOCUS       AX346822               12590 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 1893 from Patent WO0200928.
ACCESSION   AX346822
```

```
VERSION     AX346822.1  GI:18494708
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 1893 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .12590
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT     2516 a    310 c   3233 g   6531 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 6;  Length 12590;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2363 TCATCTACTAAATAAATAAAAA 2384
              ||||||||||||||||||||||
Db       5900 TCATCTACTAAATAAATAAAAA 5879


RESULT 22
AP006134
LOCUS       AP006134               21590 bp    DNA     linear   PLN 19-MAR-2003
DEFINITION  Lotus japonicus genomic DNA, chromosome 1, clone:LjT09O08, TM0231b,
            complete sequence.
ACCESSION   AP006134
VERSION     AP006134.1  GI:29122771
KEYWORDS    HTG.
SOURCE      Lotus japonicus
  ORGANISM  Lotus japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Kato,T., Sato,S., Nakamura,Y. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. III. Sequence
            Features and Mapping of Sixty-two TAC Clones Which Cover the 6.7 Mb
            Regions of the Genome
  JOURNAL   DNA Res. 10, 27-33 (2003)
REFERENCE   2  (bases 1 to 21590)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2002) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1. .21590
                     /organism="Lotus japonicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:34305"
                     /chromosome="1"
                     /clone="LjT09O08"
                     /clone_lib="LjT library"
                     /note="TM0231b, a part of TAC clone:TM0231."
BASE COUNT     6559 a   3827 c   4386 g   6818 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 8;  Length 21590;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8617 TAATATACCAATAATAGACTTA 8638
              ||||||||||||||||||||||
Db       8897 TAATATACCAATAATAGACTTA 8918


RESULT 23
AC139177/c
LOCUS       AC139177               41951 bp    DNA     linear   HTG 25-JAN-2003
DEFINITION  Ciona savignyi clone -1L23, WORKING DRAFT SEQUENCE, 5 unordered
            pieces.
ACCESSION   AC139177
VERSION     AC139177.1  GI:27901883
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Ciona savignyi
  ORGANISM  Ciona savignyi
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 41951)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Ciona savignyi, clone -1L23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 41951)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JAN-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L18113
                Center clone name: 1_L_23
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 40698 bases at least Q40
                Consensus quality: 41388 bases at least Q30
                Consensus quality: 41487 bases at least Q20
                Insert size: 42000; agarose-fp
                Insert size: 41551; sum-of-contigs
                Quality coverage: 13.0 in Q20 bases; agarose-fp
                Quality coverage: 13.1 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
```

```
            * as soon as it is available and the accession number will
            * be preserved.
            *         1       12: contig of 12 bp in length
            *        13      112: gap of 100 bp
            *       113      498: contig of 386 bp in length
            *       499      598: gap of 100 bp
            *       599     1785: contig of 1187 bp in length
            *      1786     1885: gap of 100 bp
            *      1886    30433: contig of 28548 bp in length
            *     30434    30533: gap of 100 bp
            *     30534    41951: contig of 11418 bp in length.
FEATURES             Location/Qualifiers
     source          1. .41951
                     /organism="Ciona savignyi"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:51511"
                     /clone="-1L23"
                     /clone_lib="CHORI-301 Ciona BAC6"
     misc_feature    1. .12
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    113. .498
                     /note="assembly_fragment"
     misc_feature    599. .1785
                     /note="assembly_fragment"
     misc_feature    1886. .30433
                     /note="assembly_fragment"
     misc_feature    30534. .41951
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    12063 a    8284 c    9133 g   12071 t     400 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 41951;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1356 TTTAAATGGAAAAACAAAAAAA 1377
              ||||||||||||||||||||||
Db      20876 TTTAAATGGAAAAACAAAAAAA 20855


RESULT 24
AC021334/c
LOCUS       AC021334               77322 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens clone RP11-24F23, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC021334
VERSION     AC021334.2  GI:9130211
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 77322)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-24F23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 77322)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6705772.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4447
                Center clone name: 24_F_23
            --------------.
            * NOTE: This record contains 86 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *         1       805: contig of 805 bp in length
            *       806       905: gap of 100 bp
            *       906      1704: contig of 799 bp in length
            *      1705      1804: gap of 100 bp
            *      1805      2598: contig of 794 bp in length
            *      2599      2698: gap of 100 bp
            *      2699      3485: contig of 787 bp in length
            *      3486      3585: gap of 100 bp
            *      3586      4394: contig of 809 bp in length
            *      4395      4494: gap of 100 bp
            *      4495      5302: contig of 808 bp in length
            *      5303      5402: gap of 100 bp
            *      5403      6201: contig of 799 bp in length
            *      6202      6301: gap of 100 bp
            *      6302      7117: contig of 816 bp in length
            *      7118      7217: gap of 100 bp
            *      7218      8027: contig of 810 bp in length
            *      8028      8127: gap of 100 bp
            *      8128      8900: contig of 773 bp in length
            *      8901      9000: gap of 100 bp
            *      9001      9816: contig of 816 bp in length
            *      9817      9916: gap of 100 bp
            *      9917     10714: contig of 798 bp in length
            *     10715     10814: gap of 100 bp
            *     10815     11623: contig of 809 bp in length
            *     11624     11723: gap of 100 bp
            *     11724     12531: contig of 808 bp in length
            *     12532     12631: gap of 100 bp
            *     12632     13439: contig of 808 bp in length
            *     13440     13539: gap of 100 bp
            *     13540     14343: contig of 804 bp in length
            *     14344     14443: gap of 100 bp
            *     14444     15245: contig of 802 bp in length
            *     15246     15345: gap of 100 bp
            *     15346     16134: contig of 789 bp in length
            *     16135     16234: gap of 100 bp
            *     16235     17060: contig of 826 bp in length
            *     17061     17160: gap of 100 bp
            *     17161     17939: contig of 779 bp in length
            *     17940     18039: gap of 100 bp
            *     18040     18846: contig of 807 bp in length
            *     18847     18946: gap of 100 bp
```

```
*    18947    19764: contig of 818 bp in length
*    19765    19864: gap of 100 bp
*    19865    20671: contig of 807 bp in length
*    20672    20771: gap of 100 bp
*    20772    21583: contig of 812 bp in length
*    21584    21683: gap of 100 bp
*    21684    22493: contig of 810 bp in length
*    22494    22593: gap of 100 bp
*    22594    23373: contig of 780 bp in length
*    23374    23473: gap of 100 bp
*    23474    24282: contig of 809 bp in length
*    24283    24382: gap of 100 bp
*    24383    25186: contig of 804 bp in length
*    25187    25286: gap of 100 bp
*    25287    26089: contig of 803 bp in length
*    26090    26189: gap of 100 bp
*    26190    26983: contig of 794 bp in length
*    26984    27083: gap of 100 bp
*    27084    27882: contig of 799 bp in length
*    27883    27982: gap of 100 bp
*    27983    28782: contig of 800 bp in length
*    28783    28882: gap of 100 bp
*    28883    29680: contig of 798 bp in length
*    29681    29780: gap of 100 bp
*    29781    30581: contig of 801 bp in length
*    30582    30681: gap of 100 bp
*    30682    31449: contig of 768 bp in length
*    31450    31549: gap of 100 bp
*    31550    32365: contig of 816 bp in length
*    32366    32465: gap of 100 bp
*    32466    33285: contig of 820 bp in length
*    33286    33385: gap of 100 bp
*    33386    34183: contig of 798 bp in length
*    34184    34283: gap of 100 bp
*    34284    35082: contig of 799 bp in length
*    35083    35182: gap of 100 bp
*    35183    35960: contig of 778 bp in length
*    35961    36060: gap of 100 bp
*    36061    36831: contig of 771 bp in length
*    36832    36931: gap of 100 bp
*    36932    37748: contig of 817 bp in length
*    37749    37848: gap of 100 bp
*    37849    38651: contig of 803 bp in length
*    38652    38751: gap of 100 bp
*    38752    39572: contig of 821 bp in length
*    39573    39672: gap of 100 bp
*    39673    40445: contig of 773 bp in length
*    40446    40545: gap of 100 bp
*    40546    41357: contig of 812 bp in length
*    41358    41457: gap of 100 bp
*    41458    42262: contig of 805 bp in length
*    42263    42362: gap of 100 bp
*    42363    43167: contig of 805 bp in length
*    43168    43267: gap of 100 bp
*    43268    44075: contig of 808 bp in length
*    44076    44175: gap of 100 bp
*    44176    44967: contig of 792 bp in length
*    44968    45067: gap of 100 bp
*    45068    45876: contig of 809 bp in length
*    45877    45976: gap of 100 bp
*    45977    46758: contig of 782 bp in length
*    46759    46858: gap of 100 bp
*    46859    47649: contig of 791 bp in length
*    47650    47749: gap of 100 bp
*    47750    48547: contig of 798 bp in length
*    48548    48647: gap of 100 bp
*    48648    49434: contig of 787 bp in length
*    49435    49534: gap of 100 bp
*    49535    50342: contig of 808 bp in length
*    50343    50442: gap of 100 bp
*    50443    51257: contig of 815 bp in length
*    51258    51357: gap of 100 bp
*    51358    52167: contig of 810 bp in length
*    52168    52267: gap of 100 bp
*    52268    53074: contig of 807 bp in length
*    53075    53174: gap of 100 bp
*    53175    53973: contig of 799 bp in length
*    53974    54073: gap of 100 bp
*    54074    54866: contig of 793 bp in length
*    54867    54966: gap of 100 bp
*    54967    55760: contig of 794 bp in length
*    55761    55860: gap of 100 bp
*    55861    56650: contig of 790 bp in length
*    56651    56750: gap of 100 bp
*    56751    57540: contig of 790 bp in length
*    57541    57640: gap of 100 bp
*    57641    58430: contig of 790 bp in length
*    58431    58530: gap of 100 bp
*    58531    59313: contig of 783 bp in length
*    59314    59413: gap of 100 bp
*    59414    60237: contig of 824 bp in length
*    60238    60337: gap of 100 bp
*    60338    61129: contig of 792 bp in length
*    61130    61229: gap of 100 bp
*    61230    62043: contig of 814 bp in length
*    62044    62143: gap of 100 bp
*    62144    62951: contig of 808 bp in length
*    62952    63051: gap of 100 bp
*    63052    63844: contig of 793 bp in length
*    63845    63944: gap of 100 bp

  Query Match             0.2%;  Score 22;  DB 2;  Length 77322;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8003 AAGATGGTTCCTGACAATTATT 8024
             ||||||||||||||||||||||
Db      32732 AAGATGGTTCCTGACAATTATT 32711


RESULT 25
AC108356_2/c
WPCOMMENT
  Sequence split into 4 fragments  LOCUS AC108356 Accession AC108356
     Fragment Name      Begin     End
     AC108356_0             1    110000
     AC108356_1        100001    210000
     AC108356_2        200001    310000
     AC108356_3        300001    392844
Continuation (3 of 4) of AC108356 from base 200001 (AC108356 Rattus norvegicus clone

  Query Match             0.2%;  Score 22;  DB 2;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3928 TTTAAAAATATTTTTGAGGCTT 3949
             ||||||||||||||||||||||
Db      34954 TTTAAAAATATTTTTGAGGCTT 34933


RESULT 26
AC132894
LOCUS       AC132894              146761 bp    DNA     linear   HTG 13-MAR-2003
DEFINITION  Mus musculus clone RP24-275D6, WORKING DRAFT SEQUENCE, 7 unordered
            pieces.
ACCESSION   AC132894
VERSION     AC132894.3  GI:28933844
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 146761)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-275D6
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 146761)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 146761)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 13, 2003 this sequence version replaced gi:28412055.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L27149
                Center clone name: 275_D_6
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 145475 bases at least Q40
                Consensus quality: 145812 bases at least Q30
                Consensus quality: 146015 bases at least Q20
                Insert size: 145000; agarose-fp
                Insert size: 146161; sum-of-contigs
                Quality coverage: 9.9 in Q20 bases; agarose-fp
                Quality coverage: 9.8 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    44249: contig of 44249 bp in length
            *    44250    44349: gap of 100 bp
            *    44350    47768: contig of 3419 bp in length
            *    47769    47868: gap of 100 bp
            *    47869    53239: contig of 5371 bp in length
            *    53240    53339: gap of 100 bp
            *    53340    60564: contig of 7225 bp in length
            *    60565    60664: gap of 100 bp
            *    60665    67416: contig of 6752 bp in length
            *    67417    67516: gap of 100 bp
            *    67517   114331: contig of 46815 bp in length
            *   114332   114431: gap of 100 bp
            *   114432   146761: contig of 32330 bp in length.
FEATURES             Location/Qualifiers
     source          1. .146761
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-275D6"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .44249
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    44350. .47768
                     /note="assembly_fragment"
     misc_feature    47869. .53239
                     /note="assembly_fragment"
     misc_feature    53340. .60564
                     /note="assembly_fragment"
     misc_feature    60665. .67416
                     /note="assembly_fragment"
     misc_feature    67517. .114331
                     /note="assembly_fragment"
     misc_feature    114432. .146761
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    43953 a  31331 c  30438 g  40439 t    600 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 146761;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2391 AATGGAAAACATTGTGTTTTTA 2412
              ||||||||||||||||||||||
Db      18263 AATGGAAAACATTGTGTTTTTA 18284


RESULT 27
AC126427
LOCUS       AC126427              151506 bp    DNA     linear   ROD 04-DEC-2002
DEFINITION  Mus musculus chromosome 15 clone RP24-365B12, complete sequence.
ACCESSION   AC126427
VERSION     AC126427.4  GI:26024144
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 151506)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151506)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (05-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 151506)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 151506)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Dec 4, 2002 this sequence version replaced gi:25189036.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information ----------------
            Center project name: M_BB0365B12
            ----------------------------------------------------.
FEATURES             Location/Qualifiers
     source          1. .151506
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="15"
                     /clone="RP24-365B12"
BASE COUNT    47035 a  29325 c  29268 g  45878 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 10;  Length 151506;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4240 TGACATTTTACTCCAAACTATT 4261
             .  ||||||||||||||||||||
Db      94590 TGACATTTTACTCCAAACTATT 94611


RESULT 28
AC135841/c
LOCUS       AC135841              155852 bp    DNA     linear   HTG 22-NOV-2002
DEFINITION  Canis familiaris clone RP81-27F9, WORKING DRAFT SEQUENCE, 6 ordered
            pieces.
ACCESSION   AC135841
VERSION     AC135841.2  GI:25167119
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 155852)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Cariaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
            Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J.,
            Paguirigan,C., Pearson,R., Portnoy,M.E., Prasad,A.,
            Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
            Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 155852)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 155852)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Nov 22, 2002 this sequence version replaced gi:24270705.
            ------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            ------------- Project Information
                 Center project name: cpe
                 Center clone name: 027F09

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            ------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 154739 bases at least Q40
                 Consensus quality: 155055 bases at least Q30
                 Consensus quality: 155254 bases at least Q20
                 Insert size: 132000; agarose-fp
                 Insert size: 155352; sum-of-contigs
                 Quality coverage: 11.16x in Q20 bases; agarose-fp
                 Quality coverage: 9.48x in Q20 bases; sum-of-contigs
            -------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    43121: contig of 43121 bp in length
            *    43122    43221: gap of unknown length
            *    43222    55187: contig of 11966 bp in length
            *    55188    55287: gap of unknown length
            *    55288   107968: contig of 52681 bp in length
            *   107969   108068: gap of unknown length
            *   108069   134900: contig of 26832 bp in length
            *   134901   135000: gap of unknown length
            *   135001   141689: contig of 6689 bp in length
            *   141690   141789: gap of unknown length
            *   141790   155852: contig of 14063 bp in length.
FEATURES             Location/Qualifiers
     source          1. .155852
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9615"
                     /clone="RP81-27F9"
                     /clone_lib="RP81"
     misc_feature    1. .97128
                     /note="clone overlaps with GenBank Accession Number
                     AC136889 clone RP81-234L23 (center project name cpd)"
     misc_feature    1. .43121
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    43222. .55187
                     /note="assembly_fragment"
```

```
     misc_feature    55288. .107968
                     /note="assembly_fragment"
     misc_feature    108069. .134900
                     /note="assembly_fragment"
     misc_feature    128481. .155852
                     /note="clone overlaps with GenBank Accession Number
                     AC136888 clone RP81-209D19 (center project name cpf)"
     misc_feature    135001. .141689
                     /note="assembly_fragment"
     misc_feature    141790. .155852
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    40315 a  35191 c  34751 g  45094 t    501 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 155852;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2371 TAAATAAATAAAAACAAGCAAA 2392
              ||||||||||||||||||||||
Db      56681 TAAATAAATAAAAACAAGCAAA 56660


RESULT 29
AC136889/c
LOCUS       AC136889            167615 bp    DNA     linear   HTG 30-NOV-2002
DEFINITION  Canis familiaris clone RP81-234L23, WORKING DRAFT SEQUENCE, 6
            ordered pieces.
ACCESSION   AC136889
VERSION     AC136889.2  GI:25956350
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 167615)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Cariaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
            Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J.,
            Paguirigan,C., Pearson,R., Portnoy,M.E., Prasad,A.,
            Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
            Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167615)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 167615)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Nov 30, 2002 this sequence version replaced gi:24819765.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: cpd
                 Center clone name: 234L23

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 166369 bases at least Q40
                 Consensus quality: 166672 bases at least Q30
                 Consensus quality: 166969 bases at least Q20
                 Insert size: 148000; agarose-fp
                 Insert size: 167115; sum-of-contigs
                 Quality coverage: 9.79x in Q20 bases; agarose-fp
                 Quality coverage: 8.67x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    44852: contig of 44852 bp in length
            *    44853    44952: gap of unknown length
            *    44953    66373: contig of 21421 bp in length
            *    66374    66473: gap of unknown length
            *    66474   114683: contig of 48210 bp in length
            *   114684   114783: gap of unknown length
            *   114784   124213: contig of 9430 bp in length
            *   124214   124313: gap of unknown length
            *   124314   134840: contig of 10527 bp in length
            *   134841   134940: gap of unknown length
            *   134941   167615: contig of 32675 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167615
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9615"
                     /clone="RP81-234L23"
                     /clone_lib="RP81"
     misc_feature    1. .96846
                     /note="clone overlaps with GenBank Accession Number
                     AC137660 clone RP81-422C13 (center project name cpc)"
     misc_feature    1. .44852
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    44953. .66373
                     /note="assembly_fragment"
     misc_feature    66474. .114683
                     /note="assembly_fragment"
     misc_feature    71846. .167615
                     /note="clone overlaps with GenBank Accession Number
                     AC135841 clone RP81-27F9 (center project name cpe)"
     misc_feature    114784. .124213
                     /note="assembly_fragment"
     misc_feature    124314. .134840
                     /note="assembly_fragment"
     misc_feature    134941. .167615
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    44693 a  34653 c  34686 g  53083 t    500 others
ORIGIN
```

```
  Query Match          0.2%;  Score 22;  DB 2;  Length 167615;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2371 TAAATAAATAAAAACAAGCAAA 2392
              ||||||||||||||||||||||
Db     127158 TAAATAAATAAAAACAAGCAAA 127137


RESULT 30
AC023203
LOCUS       AC023203               167910 bp    DNA     linear   HTG 12-MAR-2000
DEFINITION  Homo sapiens chromosome 8 clone RP11-352L22 map 8, WORKING DRAFT
            SEQUENCE, 21 unordered pieces.
ACCESSION   AC023203
VERSION     AC023203.2  GI:7229857
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167910)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-352L22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167910)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 12, 2000 this sequence version replaced gi:6957767.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6195
                Center clone name: 352_L_22
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 157035 bases at least Q40
                Consensus quality: 162186 bases at least Q30
                Consensus quality: 164379 bases at least Q20
                Insert size: 182000; agarose-fp
                Insert size: 165910; sum-of-contigs
                Quality coverage: 4.3 in Q20 bases; agarose-fp
                Quality coverage: 4.7 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1313: contig of 1313 bp in length
            *    1314     1413: gap of 100 bp
            *    1414     2425: contig of 1012 bp in length
            *    2426     2525: gap of 100 bp
            *    2526     3678: contig of 1153 bp in length
            *    3679     3778: gap of 100 bp
            *    3779     5347: contig of 1569 bp in length
            *    5348     5447: gap of 100 bp
            *    5448     7468: contig of 2021 bp in length
            *    7469     7568: gap of 100 bp
            *    7569     9316: contig of 1748 bp in length
            *    9317     9416: gap of 100 bp
            *    9417    10057: contig of 641 bp in length
            *   10058    10157: gap of 100 bp
            *   10158    12333: contig of 2176 bp in length
            *   12334    12433: gap of 100 bp
            *   12434    13207: contig of 774 bp in length
            *   13208    13307: gap of 100 bp
            *   13308    16436: contig of 3129 bp in length
            *   16437    16536: gap of 100 bp
            *   16537    22107: contig of 5571 bp in length
            *   22108    22207: gap of 100 bp
            *   22208    28325: contig of 6118 bp in length
            *   28326    28425: gap of 100 bp
            *   28426    35027: contig of 6602 bp in length
            *   35028    35127: gap of 100 bp
            *   35128    45706: contig of 10579 bp in length
            *   45707    45806: gap of 100 bp
            *   45807    58680: contig of 12874 bp in length
            *   58681    58780: gap of 100 bp
            *   58781    74137: contig of 15357 bp in length
            *   74138    74237: gap of 100 bp
            *   74238    84682: contig of 10445 bp in length
            *   84683    84782: gap of 100 bp
            *   84783   100921: contig of 16139 bp in length
            *  100922   101021: gap of 100 bp
            *  101022   119956: contig of 18935 bp in length
            *  119957   120056: gap of 100 bp
            *  120057   140414: contig of 20358 bp in length
            *  140415   140514: gap of 100 bp
            *  140515   167910: contig of 27396 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167910
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-352L22"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1313
                     /note="assembly_fragment"
     misc_feature    1414. .2425
                     /note="assembly_fragment"
     misc_feature    2526. .3678
                     /note="assembly_fragment"
     misc_feature    3779. .5347
                     /note="assembly_fragment"
     misc_feature    5448. .7468
                     /note="assembly_fragment"
     misc_feature    7569. .9316
                     /note="assembly_fragment"
     misc_feature    9417. .10057
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    10158. .12333
                     /note="assembly_fragment"
```

```
     misc_feature    12434. .13207
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    13308. .16436
                     /note="assembly_fragment"
     misc_feature    16537. .22107
                     /note="assembly_fragment"
     misc_feature    22208. .28325
                     /note="assembly_fragment"
     misc_feature    28426. .35027
                     /note="assembly_fragment"
     misc_feature    35128. .45706
                     /note="assembly_fragment"
     misc_feature    45807. .58680
                     /note="assembly_fragment"
     misc_feature    58781. .74137
                     /note="assembly_fragment"
     misc_feature    74238. .84682
                     /note="assembly_fragment"
     misc_feature    84783. .100921
                     /note="assembly_fragment"
     misc_feature    101022. .119956
                     /note="assembly_fragment"
     misc_feature    120057. .140414
                     /note="assembly_fragment"
     misc_feature    140515. .167910
                     /note="assembly_fragment"
BASE COUNT    55319 a  29761 c  28884 g  51943 t   2003 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 167910;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7291 CTTGTACATATGATTTAAAAGA 7312
              ||||||||||||||||||||||
Db      79382 CTTGTACATATGATTTAAAAGA 79403


RESULT 31
AC023035
LOCUS       AC023035              169159 bp    DNA     linear   HTG 03-MAR-2001
DEFINITION  Homo sapiens chromosome 7 clone RP11-515B13 map 7, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC023035
VERSION     AC023035.4  GI:13184191
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169159)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 7, clone RP11-515B13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169159)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 2, 2001 this sequence version replaced gi:8076882.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6695
                Center clone name: 515_B_13
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 95% of reads
                Sequencing vector: Plasmid; n/a; 5% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 159721 bases at least Q40
                Consensus quality: 164482 bases at least Q30
                Consensus quality: 166441 bases at least Q20
                Insert size: 173000; agarose-fp
                Insert size: 167859; sum-of-contigs
                Quality coverage: 4.5 in Q20 bases; agarose-fp
                Quality coverage: 4.6 in Q20 ba.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    23924: contig of 23924 bp in length
            *    23925    24024: gap of 100 bp
            *    24025    84038: contig of 60014 bp in length
            *    84039    84138: gap of 100 bp
            *    84139    85661: contig of 1523 bp in length
            *    85662    85761: gap of 100 bp
            *    85762    87360: contig of 1599 bp in length
            *    87361    87460: gap of 100 bp
            *    87461    89368: contig of 1908 bp in length
            *    89369    89468: gap of 100 bp
            *    89469    92414: contig of 2946 bp in length
            *    92415    92514: gap of 100 bp
            *    92515    94933: contig of 2419 bp in length
            *    94934    95033: gap of 100 bp
            *    95034    98174: contig of 3141 bp in length
            *    98175    98274: gap of 100 bp
            *    98275   102192: contig of 3918 bp in length
            *   102193   102292: gap of 100 bp
            *   102293   114914: contig of 12622 bp in length
            *   114915   115014: gap of 100 bp
            *   115015   132995: contig of 17981 bp in length
            *   132996   133095: gap of 100 bp
            *   133096   151179: contig of 18084 bp in length
            *   151180   151279: gap of 100 bp
            *   151280   168267: contig of 16988 bp in length
            *   168268   168367: gap of 100 bp
            *   168368   169159: contig of 792 bp in length.
FEATURES             Location/Qualifiers
     source          1. .169159
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7"
                     /clone="RP11-515B13"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .23924
```

```
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    24025. .84038
                     /note="assembly_fragment"
     misc_feature    84139. .85661
                     /note="assembly_fragment"
     misc_feature    85762. .87360
                     /note="assembly_fragment"
     misc_feature    87461. .89368
                     /note="assembly_fragment"
     misc_feature    89469. .92414
                     /note="assembly_fragment"
     misc_feature    92515. .94933
                     /note="assembly_fragment"
     misc_feature    95034. .98174
                     /note="assembly_fragment"
     misc_feature    98275. .102192
                     /note="assembly_fragment"
     misc_feature    102293. .114914
                     /note="assembly_fragment"
     misc_feature    115015. .132995
                     /note="assembly_fragment"
     misc_feature    133096. .151179
                     /note="assembly_fragment"
     misc_feature    151280. .168267
                     /note="assembly_fragment"
     misc_feature    168368. .169159
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    56803 a  28948 c  30010 g  52096 t   1302 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 169159;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7291 CTTGTACATATGATTTAAAAGA 7312
              ||||||||||||||||||||||
Db     165549 CTTGTACATATGATTTAAAAGA 165570


RESULT 32
AC126437/c
LOCUS       AC126437              169332 bp    DNA     linear   HTG 14-APR-2003
DEFINITION  Mus musculus chromosome UNK clone RP24-211B22, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC126437
VERSION     AC126437.2  GI:29611697
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 169332)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169332)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 169332)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-APR-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Apr 8, 2003 this sequence version replaced gi:21699705.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------- Project Information ----------
            Center project name: M_BB0211B22
            ----------------------------------------------------
            ---------- Summary Statistics  ----------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 167467 bases at least Q40
            Consensus quality: 167669 bases at least Q30
            Consensus quality: 167760 bases at least Q20
            ----------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    46975: contig of 46975 bp in length
            *    46976    47075: gap of unknown length
            *    47076   169332: contig of 122257 bp in length.
FEATURES             Location/Qualifiers
     source          1. .169332
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP24-211B22"
     misc_feature    1. .46975
                     /note="assembly_name:Contig20"
     misc_feature    47076. .169332
                     /note="assembly_name:Contig21"
BASE COUNT    54361 a  31199 c  31049 g  52622 t    101 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 169332;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5159 CTTGTAAAATATCATCCAGTGA 5180
              ||||||||||||||||||||||
Db      99552 CTTGTAAAATATCATCCAGTGA 99531


RESULT 33
AL929503
LOCUS       AL929503              174012 bp    DNA     linear   VRT 22-JAN-2003
DEFINITION  Zebrafish DNA sequence from clone CH211-158I6, complete sequence.
ACCESSION   AL929503
VERSION     AL929503.4  GI:27652799
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 174012)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 11, 2003 this sequence version replaced gi:26985042.
            ---------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
```

```
          Contact: zfish-help@sanger.ac.uk
          --------------
          During sequence assembly data is compared from overlapping clones.
          Where differences are found these are annotated as variations
          together with a note of the overlapping clone name. Note that the
          variation annotation may not be found in the sequence submission
          corresponding to the overlapping clone, as we submit sequences with
          only a small overlap as described above.
          This sequence was finished as follows unless otherwise noted: all
          regions were either double-stranded or sequenced with an alternate
          chemistry or covered by high quality data (i.e., phred quality >=
          30); an attempt was made to resolve all sequencing problems, such
          as compressions and repeats; all regions were covered by at least
          one plasmid subclone or more than one M13 subclone; and the
          assembly was confirmed by restriction digest, except on the rare
          occasion of the clone being a YAC.
          The following abbreviations are used to associate primary accession
          numbers given in the feature table with their source databases:
          Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
          on the WORMPEP database can be found at
          http://www.sanger.ac.uk/Projects/C_elegans/wormpep Repeat names
          beginning 'Dr' were identified by the Recon repeat discovery system
          (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr'
          were identified by Rick Waterman (Stephen Johnson lab, WashU). For
          further information see http://www/Projects/D_rerio/fishmask.shtml
          CH211-158I6 is from a CHORI-211 BAC library
          VECTOR: pTARBAC2.1.
FEATURES             Location/Qualifiers
     source          1. .174012
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-158I6"
                     /clone_lib="CHORI-211"
BASE COUNT    57710 a  31631 c  30657 g  54014 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 5;  Length 174012;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6422 TTAAACTTTTTTTATTCAAGAT 6443
             ||||||||||||||||||||||
Db     123562 TTAAACTTTTTTTATTCAAGAT 123583


RESULT 34
AC025888
LOCUS       AC025888              178201 bp    DNA     linear   PRI 28-JUN-2002
DEFINITION  Homo sapiens chromosome 18, clone RP11-479H18, complete sequence.
ACCESSION   AC025888
VERSION     AC025888.8  GI:16931021
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 178201)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-479H18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178201)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 178201)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 15, 2001 this sequence version replaced gi:15431162.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L8234
                Center clone name: 479_H_18
            --------------.
FEATURES             Location/Qualifiers
     source          1. .178201
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-479H18"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   386. .424
                     /rpt_family="AT_rich"
     repeat_region   2196. .4358
                     /rpt_family="L1MA4"
     repeat_region   4359. .5242
                     /rpt_family="L1MA6"
     repeat_region   5236. .5712
                     /rpt_family="L1MA4"
     repeat_region   complement(5713. .6020)
                     /rpt_family="AluY"
     repeat_region   6021. .7422
                     /rpt_family="L1MA4"
```

```
repeat_region   7423. .7722
                /rpt_family="AluSp"
repeat_region   7723. .10671
                /rpt_family="L1MA4"
repeat_region   11300. .11745
                /rpt_family="L3"
repeat_region   12067. .12087
                /rpt_family="AT_rich"
repeat_region   complement(12276. .12344)
                /rpt_family="MIR"
repeat_region   12512. .12551
                /rpt_family="AT_rich"
repeat_region   complement(12660. .12766)
                /rpt_family="AluSq/x"
repeat_region   complement(12962. .13355)
                /rpt_family="L1ME4A"
unsure          complement(13100. .13124)
                /note="<30 qual SNGL region"
repeat_region   13365. .13424
                /rpt_family="(TCTA)n"
repeat_region   complement(13544. .14025)
                /rpt_family="L2"
repeat_region   14026. .14070
                /rpt_family="AT_rich"
repeat_region   complement(14076. .14147)
                /rpt_family="L1ME3A"
repeat_region   complement(14356. .15045)
                /rpt_family="L2"
repeat_region   15245. .15425
                /rpt_family="L1MA1"
repeat_region   complement(15454. .15544)
                /rpt_family="HAL1b"
repeat_region   15545. .15837
                /rpt_family="AluSq"
repeat_region   complement(15838. .15980)
                /rpt_family="HAL1b"
repeat_region   17168. .17344
                /rpt_family="MLT1B"
repeat_region   17345. .20294
                /rpt_family="L1PA13"
repeat_region   20295. .20539
                /rpt_family="MLT1B"
repeat_region   22049. .22341
                /rpt_family="AluJb"
repeat_region   22481. .22586
                /rpt_family="L1PA5"
repeat_region   22587. .22640
                /rpt_family="AT_rich"
repeat_region   complement(22906. .23147)
                /rpt_family="MIR"
repeat_region   complement(23767. .23995)
                /rpt_family="L1M4"
repeat_region   24391. .24692
                /rpt_family="AluSg"
repeat_region   25142. .25449
                /rpt_family="AluY"
repeat_region   25460. .25555
                /rpt_family="MIR"
repeat_region   25733. .25764
                /rpt_family="AT_rich"
repeat_region   26135. .26164
                /rpt_family="AT_rich"
repeat_region   26233. .26318
                /rpt_family="(TA)n"
repeat_region   26415. .26543
                /rpt_family="MER103"
repeat_region   26544. .26602
                /rpt_family="(CATATA)n"
repeat_region   complement(26620. .26829)
                /rpt_family="MER20"
repeat_region   26830. .26874
                /rpt_family="(CA)n"
repeat_region   26894. .26917
                /rpt_family="(CA)n"
repeat_region   27371. .27398
                /rpt_family="(TG)n"
repeat_region   27427. .27463
                /rpt_family="(TG)n"
repeat_region   complement(28058. .28466)
                /rpt_family="L1M4"
repeat_region   28694. .28878
                /rpt_family="MIR3"
unsure          complement(29200. .29218)
                /note="<30 qual SNGL region"
unsure          complement(29226. .29245)
                /note="<30 qual SNGL region"
repeat_region   29454. .29483
                /rpt_family="AT_rich"
repeat_region   29630. .29680
                /rpt_family="AT_rich"
repeat_region   30154. .30176
                /rpt_family="AT_rich"
repeat_region   complement(31185. .31459)
                /rpt_family="AluSx"
repeat_region   complement(31677. .31810)
                /rpt_family="MIR"
repeat_region   33321. .33357
                /rpt_family="AT_rich"
repeat_region   33594. .33669
                /rpt_family="L1ME"
repeat_region   34795. .34815
                /rpt_family="AT_rich"
```

Query Match 0.2%; Score 22; DB 9; Length 178201;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       7291 CTTGTACATATGATTTAAAAGA 7312
              ||||||||||||||||||||||
Db     100937 CTTGTACATATGATTTAAAAGA 100958
```

RESULT 35
AC120424

```
LOCUS       AC120424              180213 bp    DNA     linear   HTG 12-MAR-2003
DEFINITION  Mus musculus clone RP24-498P8, WORKING DRAFT SEQUENCE, 15 unordered
            pieces.
ACCESSION   AC120424
VERSION     AC120424.3  GI:28927851
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 180213)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-498P8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180213)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
```

```
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 180213)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 12, 2003 this sequence version replaced gi:28626861.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L26040
                Center clone name: 498_P_8
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 175503 bases at least Q40
                Consensus quality: 177104 bases at least Q30
                Consensus quality: 177893 bases at least Q20
                Insert size: 170000; agarose-fp
                Insert size: 178813; sum-of-contigs
                Quality coverage: 8.2 in Q20 bases; agarose-fp
                Quality coverage: 7.8 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      568: contig of 568 bp in length
            *      569      668: gap of 100 bp
            *      669     1738: contig of 1070 bp in length
            *     1739     1838: gap of 100 bp
            *     1839     3580: contig of 1742 bp in length
            *     3581     3680: gap of 100 bp
            *     3681     6817: contig of 3137 bp in length
            *     6818     6917: gap of 100 bp
            *     6918     9403: contig of 2486 bp in length
            *     9404     9503: gap of 100 bp
            *     9504    11809: contig of 2306 bp in length
            *    11810    11909: gap of 100 bp
            *    11910    18098: contig of 6189 bp in length
            *    18099    18198: gap of 100 bp
            *    18199    24742: contig of 6544 bp in length
            *    24743    24842: gap of 100 bp
            *    24843    34037: contig of 9195 bp in length
            *    34038    34137: gap of 100 bp
            *    34138    43481: contig of 9344 bp in length
            *    43482    43581: gap of 100 bp
            *    43582    57616: contig of 14035 bp in length
            *    57617    57716: gap of 100 bp
            *    57717    74276: contig of 16560 bp in length
            *    74277    74376: gap of 100 bp
            *    74377    94469: contig of 20093 bp in length
            *    94470    94569: gap of 100 bp
            *    94570   115760: contig of 21191 bp in length
            *   115761   115860: gap of 100 bp
            *   115861   180213: contig of 64353 bp in length.
FEATURES             Location/Qualifiers
     source          1. .180213
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-498P8"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .568
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    669. .1738
                     /note="assembly_fragment"
     misc_feature    1839. .3580
                     /note="assembly_fragment"
     misc_feature    3681. .6817
                     /note="assembly_fragment"
     misc_feature    6918. .9403
                     /note="assembly_fragment"
     misc_feature    9504. .11809
                     /note="assembly_fragment"
     misc_feature    11910. .18098
                     /note="assembly_fragment"
     misc_feature    18199. .24742
                     /note="assembly_fragment"
     misc_feature    24843. .34037
                     /note="assembly_fragment"
     misc_feature    34138. .43481
                     /note="assembly_fragment"
     misc_feature    43582. .57616
                     /note="assembly_fragment"
     misc_feature    57717. .74276
                     /note="assembly_fragment"
     misc_feature    74377. .94469
                     /note="assembly_fragment"
     misc_feature    94570. .115760
                     /note="assembly_fragment"
     misc_feature    115861. .180213
                     /note="assembly_fragment"
BASE COUNT    48118 a  43188 c  43121 g  44350 t   1436 others
ORIGIN

  Query Match           0.2%;  Score 22;  DB 2;  Length 180213;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3921 CTATCTCTTTAAAAATATTTTT 3942
              ||||||||||||||||||||||
Db      58606 CTATCTCTTTAAAAATATTTTT 58627


RESULT 36
AC137678
```

```
LOCUS       AC137678             187559 bp    DNA     linear   HTG 11-MAR-2003
DEFINITION  Mus musculus clone RP24-484G16, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   AC137678
VERSION     AC137678.3  GI:28913198
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 187559)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-484G16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 187559)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-NOV-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 187559)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagópian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 11, 2003 this sequence version replaced gi:28209776.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L26700
                Center clone name: 484_G_16
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 185488 bases at least Q40
                Consensus quality: 186072 bases at least Q30
                Consensus quality: 186362 bases at least Q20
                Insert size: 164000; agarose-fp
                Insert size: 186759; sum-of-contigs
                Quality coverage: 13:7 in Q20 bases; agarose-fp
                Quality coverage: 12.0 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      791: contig of 791 bp in length
            *      792      891: gap of 100 bp
            *      892     1432: contig of 541 bp in length
            *     1433     1532: gap of 100 bp
            *     1533     2388: contig of 856 bp in length
            *     2389     2488: gap of 100 bp
            *     2489     4685: contig of 2197 bp in length
            *     4686     4785: gap of 100 bp
            *     4786    88902: contig of 84117 bp in length
            *    88903    89002: gap of 100 bp
            *    89003   104147: contig of 15145 bp in length
            *   104148   104247: gap of 100 bp
            *   104248   132010: contig of 27763 bp in length
            *   132011   132110: gap of 100 bp
            *   132111   180108: contig of 47998 bp in length
            *   180109   180208: gap of 100 bp
            *   180209   187559: contig of 7351 bp in length.
FEATURES             Location/Qualifiers
     source          1. .187559
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-484G16"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .791
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    892. .1432
                     /note="assembly_fragment"
     misc_feature    1533. .2388
                     /note="assembly_fragment"
     misc_feature    2489. .4685
                     /note="assembly_fragment"
     misc_feature    4786. .88902
                     /note="assembly_fragment"
     misc_feature    89003. .104147
                     /note="assembly_fragment"
     misc_feature    104248. .132010
                     /note="assembly_fragment"
     misc_feature    132111. .180108
                     /note="assembly_fragment"
     misc_feature    180209. .187559
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    45660 a  46248 c  47906 g  46941 t    804 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 187559;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3921 CTATCTCTTTAAAAATATTTTT 3942
```

```
                    ||||||||||||||||||||
Db     104834 CTATCTCTTTAAAAATATTTTT 104855


RESULT 37
AC126269/c
LOCUS       AC126269              187685 bp    DNA     linear   HTG 04-JUL-2002
DEFINITION  Mus musculus chromosome UNK clone RP23-230D6, WORKING DRAFT
            SEQUENCE, 18 unordered pieces.
ACCESSION   AC126269
VERSION     AC126269.1  GI:21693985
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 187685)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 187685)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT
            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------- Project Information ----------
            Center project name: M_BA0230D06
            ----------------------------------------------------
            ---------- Summary Statistics ----------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 177902 bases at least Q40
            Consensus quality: 179076 bases at least Q30
            Consensus quality: 179866 bases at least Q20
            ----------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1087: contig of 1087 bp in length
            *     1088     1187: gap of unknown length
            *     1188     2766: contig of 1579 bp in length
            *     2767     2866: gap of unknown length
            *     2867     4088: contig of 1222 bp in length
            *     4089     4188: gap of unknown length
            *     4189     6047: contig of 1859 bp in length
            *     6048     6147: gap of unknown length
            *     6148    10021: contig of 3874 bp in length
            *    10022    10121: gap of unknown length
            *    10122    12218: contig of 2097 bp in length
            *    12219    12318: gap of unknown length
            *    12319    17315: contig of 4997 bp in length
            *    17316    17415: gap of unknown length
            *    17416    20946: contig of 3531 bp in length
            *    20947    21046: gap of unknown length
            *    21047    27576: contig of 6530 bp in length
            *    27577    27676: gap of unknown length
            *    27677    34936: contig of 7260 bp in length
            *    34937    35036: gap of unknown length
            *    35037    42871: contig of 7835 bp in length
            *    42872    42971: gap of unknown length
            *    42972    56945: contig of 13974 bp in length
            *    56946    57045: gap of unknown length
            *    57046    80621: contig of 23576 bp in length
            *    80622    80721: gap of unknown length
            *    80722   100617: contig of 19896 bp in length
            *   100618   100717: gap of unknown length
            *   100718   120457: contig of 19740 bp in length
            *   120458   120557: gap of unknown length
            *   120558   139862: contig of 19305 bp in length
            *   139863   139962: gap of unknown length
            *   139963   186757: contig of 46795 bp in length
            *   186758   186857: gap of unknown length
            *   186858   187685: contig of 828 bp in length.
FEATURES             Location/Qualifiers
     source          1. .187685
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-230D6"
     misc_feature    1. .1087
                     /note="assembly_name:Contig11"
     misc_feature    1188. .2766
                     /note="assembly_name:Contig12"
     misc_feature    2867. .4088
                     /note="assembly_name:Contig13"
     misc_feature    4189. .6047
                     /note="assembly_name:Contig14"
     misc_feature    6148. .10021
                     /note="assembly_name:Contig15"
     misc_feature    10122. .12218
                     /note="assembly_name:Contig16"
     misc_feature    12319. .17315
                     /note="assembly_name:Contig17"
     misc_feature    17416. .20946
                     /note="assembly_name:Contig18"
     misc_feature    21047. .27576
                     /note="assembly_name:Contig19"
     misc_feature    27677. .34936
                     /note="assembly_name:Contig20"
     misc_feature    35037. .42871
                     /note="assembly_name:Contig21"
     misc_feature    42972. .56945
                     /note="assembly_name:Contig22"
     misc_feature    57046. .80621
                     /note="assembly_name:Contig23"
     misc_feature    80722. .100617
                     /note="assembly_name:Contig24"
     misc_feature    100718. .120457
                     /note="assembly_name:Contig25"
     misc_feature    120558. .139862
                     /note="assembly_name:Contig26"
     misc_feature    139963. .186757
                     /note="assembly_name:Contig27"
     misc_feature    186858. .187685
                     /note="assembly_name:Contig8"
BASE COUNT    50083 a  40979 c  39792 g  55111 t   1720 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 187685;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6702 AAAAAAAACAAAGATTTAAAAA 6723
              ||||||||||||||||||||||
Db      52049 AAAAAAAACAAAGATTTAAAAA 52028


RESULT 38
AC019017/c
LOCUS       AC019017              191804 bp    DNA     linear   HTG 16-MAR-2000
DEFINITION  Homo sapiens clone RP11-377H18, WORKING DRAFT SEQUENCE, 19
```

```
            unordered pieces.
ACCESSION   AC019017
VERSION     AC019017.3  GI:7249201
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 191804)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-377H18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191804)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-DEC-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 16, 2000 this sequence version replaced gi:6721245.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5524
                Center clone name: 377_H_18
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 180259 bases at least Q40
                Consensus quality: 185911 bases at least Q30
                Consensus quality: 187932 bases at least Q20
                Insert size: 165000; agarose-fp
                Insert size: 190004; sum-of-contigs
                Quality coverage: 4.8 in Q20 bases; agarose-fp
                Quality coverage: 4.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1069: contig of 1069 bp in length
            *     1070     1169: gap of 100 bp
            *     1170     2314: contig of 1145 bp in length
            *     2315     2414: gap of 100 bp
            *     2415     3920: contig of 1506 bp in length
            *     3921     4020: gap of 100 bp
            *     4021     6495: contig of 2475 bp in length
            *     6496     6595: gap of 100 bp
            *     6596     7591: contig of 996 bp in length
            *     7592     7691: gap of 100 bp
            *     7692    10537: contig of 2846 bp in length
            *    10538    10637: gap of 100 bp
            *    10638    15595: contig of 4958 bp in length
            *    15596    15695: gap of 100 bp
            *    15696    19505: contig of 3810 bp in length
            *    19506    19605: gap of 100 bp
            *    19606    26135: contig of 6530 bp in length
            *    26136    26235: gap of 100 bp
            *    26236    34763: contig of 8528 bp in length
            *    34764    34863: gap of 100 bp
            *    34864    44951: contig of 10088 bp in length
            *    44952    45051: gap of 100 bp
            *    45052    55401: contig of 10350 bp in length
            *    55402    55501: gap of 100 bp
            *    55502    67655: contig of 12154 bp in length
            *    67656    67755: gap of 100 bp
            *    67756    78045: contig of 10290 bp in length
            *    78046    78145: gap of 100 bp
            *    78146    95781: contig of 17636 bp in length
            *    95782    95881: gap of 100 bp
            *    95882   115339: contig of 19458 bp in length
            *   115340   115439: gap of 100 bp
            *   115440   134651: contig of 19212 bp in length
            *   134652   134751: gap of 100 bp
            *   134752   160095: contig of 25344 bp in length
            *   160096   160195: gap of 100 bp
            *   160196   191804: contig of 31609 bp in length.
FEATURES             Location/Qualifiers
     source          1. .191804
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-377H18"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1069
                     /note="assembly_fragment"
     misc_feature    1170. .2314
                     /note="assembly_fragment"
     misc_feature    2415. .3920
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    4021. .6495
                     /note="assembly_fragment"
     misc_feature    6596. .7591
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    7692. .10537
                     /note="assembly_fragment"
     misc_feature    10638. .15595
                     /note="assembly_fragment"
     misc_feature    15696. .19505
                     /note="assembly_fragment"
     misc_feature    19606. .26135
                     /note="assembly_fragment"
     misc_feature    26236. .34763
                     /note="assembly_fragment"
     misc_feature    34864. .44951
                     /note="assembly_fragment"
     misc_feature    45052. .55401
                     /note="assembly_fragment"
     misc_feature    55502. .67655
                     /note="assembly_fragment"
     misc_feature    67756. .78045
                     /note="assembly_fragment"
     misc_feature    78146. .95781
                     /note="assembly_fragment"
     misc_feature    95882. .115339
                     /note="assembly_fragment"
     misc_feature    115440. .134651
```

```
                     /note="assembly_fragment"
     misc_feature    134752. .160095
                     /note="assembly_fragment"
     misc_feature    160196. .191804
                     /note="assembly_fragment"
BASE COUNT    49320 a  44876 c  45615 g  50188 t   1805 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 191804;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2371 TAAATAAATAAAAACAAGCAAA 2392
             ||||||||||||||||||||||
Db      57927 TAAATAAATAAAAACAAGCAAA 57906


RESULT 39
AC084310
LOCUS       AC084310               193454 bp    DNA     linear   HTG 25-MAR-2003
DEFINITION  Mus musculus clone RP23-10C10, WORKING DRAFT SEQUENCE, 6 unordered
            pieces.
ACCESSION   AC084310
VERSION     AC084310.7  GI:29171424
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 193454)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-10C10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 193454)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 193454)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 25, 2003 this sequence version replaced gi:14971305.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5314
                Center clone name: 10_C_10
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 0% of reads
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 191659 bases at least Q40
                Consensus quality: 192389 bases at least Q30
                Consensus quality: 192785 bases at least Q20
                Insert size: 192954; sum-of-contigs
                Quality coverage: 15.1 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    62541: contig of 62541 bp in length
            *    62542    62641: gap of 100 bp
            *    62642    63673: contig of 1032 bp in length
            *    63674    63773: gap of 100 bp
            *    63774    69970: contig of 6197 bp in length
            *    69971    70070: gap of 100 bp
            *    70071    85492: contig of 15422 bp in length
            *    85493    85592: gap of 100 bp
            *    85593   150087: contig of 64495 bp in length
            *   150088   150187: gap of 100 bp
            *   150188   193454: contig of 43267 bp in length.
FEATURES             Location/Qualifiers
     source          1. .193454
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-10C10"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .62541
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    62642. .63673
                     /note="assembly_fragment"
     misc_feature    63774. .69970
                     /note="assembly_fragment"
     misc_feature    70071. .85492
                     /note="assembly_fragment"
     misc_feature    85593. .150087
                     /note="assembly_fragment"
     misc_feature    150188. .193454
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
```

```
BASE COUNT    50854 a  45452 c  46394 g  50254 t    500 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 193454;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2371 TAAATAAATAAAAACAAGCAAA 2392
              ||||||||||||||||||||||
Db      56434 TAAATAAATAAAAACAAGCAAA 56455


RESULT 40
BX005411/c
LOCUS       BX005411              194247 bp    DNA     linear   HTG 03-APR-2003
DEFINITION  Danio rerio clone DKEY-53I6, *** SEQUENCING IN PROGRESS ***, 11
            unordered pieces.
ACCESSION   BX005411
VERSION     BX005411.3  GI:29538720
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 194247)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 3, 2003 this sequence version replaced gi:27497371.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK53I6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 190103 bases at least Q40
            Consensus quality: 191299 bases at least Q30
            Consensus quality: 192128 bases at least Q20
            Insert size: 193247; sum-of-contigs
            Insert size: 178262; 6.9% error; agarose-fp
            Quality coverage: 5.72x in Q20 bases; sum-of-contigs Quality
            coverage: 6.23x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7766: contig of 7766 bp in length
            *     7767     7866: gap of 100 bp
            *     7867    16161: contig of 8295 bp in length
            *    16162    16261: gap of 100 bp
            *    16262    26497: contig of 10236 bp in length
            *    26498    26597: gap of 100 bp
            *    26598    31077: contig of 4480 bp in length
            *    31078    31177: gap of 100 bp
            *    31178    42119: contig of 10942 bp in length
            *    42120    42219: gap of 100 bp
            *    42220    76626: contig of 34407 bp in length
            *    76627    76726: gap of 100 bp
            *    76727    84911: contig of 8185 bp in length
            *    84912    85011: gap of 100 bp
            *    85012    89161: contig of 4150 bp in length
            *    89162    89261: gap of 100 bp
            *    89262    95358: contig of 6097 bp in length
            *    95359    95458: gap of 100 bp
            *    95459   160610: contig of 65152 bp in length
            *   160611   160710: gap of 100 bp
            *   160711   194247: contig of 33537 bp in length.
FEATURES             Location/Qualifiers
     source          1. .194247
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-53I6"
                     /clone_lib="DanioKey"
     misc_feature    1. .7766
                     /note="assembly_fragment:00588
                     fragment_chain:1"
     misc_feature    7867. .16161
                     /note="assembly_fragment:01956
                     fragment_chain:1"
     misc_feature    16262. .26497
                     /note="assembly_fragment:00489
                     fragment_chain:1"
     misc_feature    26598. .31077
                     /note="assembly_fragment:00687
                     fragment_chain:1"
     misc_feature    31178. .42119
                     /note="assembly_fragment:00388
                     fragment_chain:1"
     misc_feature    42220. .76626
                     /note="assembly_fragment:01832
                     fragment_chain:1"
     misc_feature    76727. .84911
                     /note="assembly_fragment:00860
                     fragment_chain:1"
     misc_feature    85012. .89161
                     /note="assembly_fragment:00424
                     fragment_chain:1"
     misc_feature    89262. .95358
                     /note="assembly_fragment:00625
                     fragment_chain:1"
     misc_feature    95459. .160610
                     /note="assembly_fragment:01177
                     fragment_chain:1"
     misc_feature    160711. .194247
                     /note="assembly_fragment:01850.0"
BASE COUNT    60470 a  35385 c  35695 g  61697 t   1000 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 194247;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2371 TAAATAAATAAAAACAAGCAAA 2392
              ||||||||||||||||||||||
Db     129367 TAAATAAATAAAAACAAGCAAA 129346


RESULT 41
AC125520/c
LOCUS       AC125520              194620 bp    DNA     linear   HTG 25-AUG-2002
DEFINITION  Mus musculus chromosome UNK clone RP24-281J17, WORKING DRAFT
            SEQUENCE, 6 unordered pieces.
ACCESSION   AC125520
VERSION     AC125520.2  GI:22476153
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 194620)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 194620)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 194620)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Aug 25, 2002 this sequence version replaced gi:21622847.

            ----------------- Genome Center ---------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information -------------------
            Center project name: M_BB0281J17
            -----------------------------------------------------
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 191909 bases at least Q40
            Consensus quality: 192395 bases at least Q30
            Consensus quality: 192694 bases at least Q20
            Insert size: 225000; agarose-fp
            Insert size: 195291; sum-of-contigs
            Quality coverage: 12.89 in Q20 bases; agarose-fp
            Quality coverage: 11.70 in Q20 bases; sum-of-contigs
            -------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     6133: contig of 6133 bp in length.
            *    6134     6233: gap of unknown length
            *    6234    28345: contig of 22112 bp in length
            *   28346    28445: gap of unknown length
            *   28446    58832: contig of 30387 bp in length
            *   58833    58932: gap of unknown length
            *   58933    95547: contig of 36615 bp in length
            *   95548    95647: gap of unknown length
            *   95648   137840: contig of 42193 bp in length
            *  137841   137940: gap of unknown length
            *  137941   194620: contig of 56680 bp in length.
FEATURES             Location/Qualifiers
     source          1. .194620
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP24-281J17"
     misc_feature    1. .6133
                     /note="assembly_name:Contig7"
     misc_feature    6234. .28345
                     /note="assembly_name:Contig8"
     misc_feature    28446. .58832
                     /note="assembly_name:Contig9"
     misc_feature    58933. .95547
                     /note="assembly_name:Contig10"
     misc_feature    95648. .137840
                     /note="assembly_name:Contig11"
     misc_feature    137941. .194620
                     /note="assembly_name:Contig12"
BASE COUNT    52598 a  44060 c  42012 g  55386 t    564 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 194620;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4713 AATTAAAAAAAATCCACATGTA 4734
             ||||||||||||||||||||||
Db     106730 AATTAAAAAAAATCCACATGTA 106709


RESULT 42
AC125195
LOCUS       AC125195              195041 bp    DNA     linear   ROD 19-DEC-2002
DEFINITION  Mus musculus chromosome 2 clone RP23-246M2, complete sequence.
ACCESSION   AC125195
VERSION     AC125195.4  GI:27261689
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 195041)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 195041)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 195041)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 195041)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-DEC-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Dec 19, 2002 this sequence version replaced gi:25046690.

            ------------------- Genome Center -------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information ----------------
            Center project name: M_BA0246M02
            -----------------------------------------------------
FEATURES             Location/Qualifiers
     source          1. .195041
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-246M2"
BASE COUNT    61244 a  35859 c  36734 g  61204 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 10;  Length 195041;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1355 CTTTAAATGGAAAAACAAAAAA 1376
             ||||||||||||||||||||||
Db      15069 CTTTAAATGGAAAAACAAAAAA 15090


RESULT 43
AC140374/c
LOCUS       AC140374              195997 bp    DNA     linear   HTG 23-FEB-2003
```

```
DEFINITION  Mus musculus chromosome UNK clone RP24-76K4, WORKING DRAFT
            SEQUENCE, 15 unordered pieces.
ACCESSION   AC140374
VERSION     AC140374.1  GI:28475632
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 195997)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 195997)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT
            ----------------- Genome Center ----------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information -------------------
            Center project name: M_BB0076K04
            ------------------------------------------------------
            ------------- Summary Statistics  --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 190208 bases at least Q40
            Consensus quality: 190804 bases at least Q30
            Consensus quality: 191344 bases at least Q20
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1103: contig of 1103 bp in length
            *    1104     1203: gap of unknown length
            *    1204     2868: contig of 1665 bp in length
            *    2869     2968: gap of unknown length
            *    2969     5952: contig of 2984 bp in length
            *    5953     6052: gap of unknown length
            *    6053     8055: contig of 2003 bp in length
            *    8056     8155: gap of unknown length
            *    8156     9710: contig of 1555 bp in length
            *    9711     9810: gap of unknown length
            *    9811    13750: contig of 3940 bp in length
            *   13751    13850: gap of unknown length
            *   13851    19283: contig of 5433 bp in length
            *   19284    19383: gap of unknown length
            *   19384    25367: contig of 5984 bp in length
            *   25368    25467: gap of unknown length
            *   25468    31246: contig of 5779 bp in length
            *   31247    31346: gap of unknown length
            *   31347    42087: contig of 10741 bp in length
            *   42088    42187: gap of unknown length
            *   42188    57959: contig of 15772 bp in length
            *   57960    58059: gap of unknown length
            *   58060    78721: contig of 20662 bp in length
            *   78722    78821: gap of unknown length
            *   78822   105763: contig of 26942 bp in length
            *  105764   105863: gap of unknown length
            *  105864   148039: contig of 42176 bp in length
            *  148040   148139: gap of unknown length
            *  148140   195997: contig of 47858 bp in length.
FEATURES             Location/Qualifiers
     source          1. .195997
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP24-76K4"
     misc_feature    1. .1103
                     /note="assembly_name:Contig9"
     misc_feature    1204. .2868
                     /note="assembly_name:Contig13"
     misc_feature    2969. .5952
                     /note="assembly_name:Contig14"
     misc_feature    6053. .8055
                     /note="assembly_name:Contig16"
     misc_feature    8156. .9710
                     /note="assembly_name:Contig17"
     misc_feature    9811. .13750
                     /note="assembly_name:Contig18"
     misc_feature    13851. .19283
                     /note="assembly_name:Contig19"
     misc_feature    19384. .25367
                     /note="assembly_name:Contig20"
     misc_feature    25468. .31246
                     /note="assembly_name:Contig21"
     misc_feature    31347. .42087
                     /note="assembly_name:Contig22"
     misc_feature    42188. .57959
                     /note="assembly_name:Contig23"
     misc_feature    58060. .78721
                     /note="assembly_name:Contig24"
     misc_feature    78822. .105763
                     /note="assembly_name:Contig25"
     misc_feature    105864. .148039
                     /note="assembly_name:Contig26"
     misc_feature    148140. .195997
                     /note="assembly_name:Contig27"
BASE COUNT    59292 a  38076 c  38564 g  58649 t   1416 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 195997;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2372 AAATAAATAAAAACAAGCAAAT 2393
              ||||||||||||||||||||||
Db     195103 AAATAAATAAAAACAAGCAAAT 195082


RESULT 44
AC102808
LOCUS       AC102808              196742 bp    DNA     linear   HTG 22-FEB-2003
DEFINITION  Mus musculus clone RP23-370M19, WORKING DRAFT SEQUENCE, 19
            unordered pieces.
ACCESSION   AC102808
VERSION     AC102808.3  GI:28467312
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 196742)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-370M19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196742)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
```

```
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 196742)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 22, 2003 this sequence version replaced gi:22004514.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L19723
                Center clone name: 370_M_19
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 192571 bases at least Q40
                Consensus quality: 193755 bases at least Q30
                Consensus quality: 194410 bases at least Q20
                Insert size: 194000; agarose-fp
                Insert size: 194942; sum-of-contigs
                Quality coverage: 9.7 in Q20 bases; agarose-fp
                Quality coverage: 9.6 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    15950: contig of 15950 bp in length
            *    15951    16050: gap of 100 bp
            *    16051    16733: contig of 683 bp in length
            *    16734    16833: gap of 100 bp
            *    16834    17517: contig of 684 bp in length
            *    17518    17617: gap of 100 bp
            *    17618    18251: contig of 634 bp in length
            *    18252    18351: gap of 100 bp
            *    18352    19023: contig of 672 bp in length
            *    19024    19123: gap of 100 bp
            *    19124    19781: contig of 658 bp in length
            *    19782    19881: gap of 100 bp
            *    19882    21250: contig of 1369 bp in length
            *    21251    21350: gap of 100 bp
            *    21351    22183: contig of 833 bp in length
            *    22184    22283: gap of 100 bp
            *    22284    22982: contig of 699 bp in length
            *    22983    23082: gap of 100 bp
            *    23083    24356: contig of 1274 bp in length
            *    24357    24456: gap of 100 bp
            *    24457    26613: contig of 2157 bp in length
            *    26614    26713: gap of 100 bp
            *    26714    33495: contig of 6782 bp in length
            *    33496    33595: gap of 100 bp
            *    33596    43856: contig of 10261 bp in length
            *    43857    43956: gap of 100 bp
            *    43957    52315: contig of 8359 bp in length
            *    52316    52415: gap of 100 bp
            *    52416    60087: contig of 7672 bp in length
            *    60088    60187: gap of 100 bp
            *    60188   126734: contig of 66547 bp in length
            *   126735   126834: gap of 100 bp
            *   126835   154518: contig of 27684 bp in length
            *   154519   154618: gap of 100 bp
            *   154619   183654: contig of 29036 bp in length
            *   183655   183754: gap of 100 bp
            *   183755   196742: contig of 12988 bp in length.
FEATURES             Location/Qualifiers
     source          1. .196742
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-370M19"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .15950
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    16051. .16733
                     /note="assembly_fragment"
     misc_feature    16834. .17517
                     /note="assembly_fragment"
     misc_feature    17618. .18251
                     /note="assembly_fragment"
     misc_feature    18352. .19023
                     /note="assembly_fragment"
     misc_feature    19124. .19781
                     /note="assembly_fragment"
     misc_feature    19882. .21250
                     /note="assembly_fragment"
     misc_feature    21351. .22183
                     /note="assembly_fragment"
     misc_feature    22284. .22982
                     /note="assembly_fragment"
     misc_feature    23083. .24356
                     /note="assembly_fragment"
     misc_feature    24457. .26613
                     /note="assembly_fragment"
     misc_feature    26714. .33495
                     /note="assembly_fragment"
     misc_feature    33596. .43856
```

```
                     /note="assembly_fragment"
     misc_feature    43957. .52315
                     /note="assembly_fragment"
     misc_feature    52416. .60087
                     /note="assembly_fragment"
     misc_feature    60188. .126734
                     /note="assembly_fragment"
     misc_feature    126835. .154518
                     /note="assembly_fragment"
     misc_feature    154619. .183654
                     /note="assembly_fragment"
     misc_feature    183755. .196742
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    59484 a  38484 c  38854 g  58119 t   1801 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 196742;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2371 TAAATAAATAAAAACAAGCAAA 2392
              ||||||||||||||||||||||
Db       2414 TAAATAAATAAAAACAAGCAAA 2435


RESULT 45
AC140314
LOCUS       AC140314               198298 bp    DNA     linear   HTG 23-FEB-2003
DEFINITION  Mus musculus chromosome UNK clone RP23-395L3, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC140314
VERSION     AC140314.1  GI:28475564
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 198298)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 198298)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT
            ----------------- Genome Center ----------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information -------------------
            Center project name: M_BA0395L03
            ------------------------------------------------------
            -------------- Summary Statistics  -------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 192267 bases at least Q40
            Consensus quality: 192737 bases at least Q30
            Consensus quality: 193088 bases at least Q20
            ------------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1432: contig of 1432 bp in length
            *     1433     1532: gap of unknown length
            *     1533     4137: contig of 2605 bp in length
            *     4138     4237: gap of unknown length
            *     4238     7586: contig of 3349 bp in length
            *     7587     7686: gap of unknown length
            *     7687    10403: contig of 2717 bp in length
            *    10404    10503: gap of unknown length
            *    10504    21007: contig of 10504 bp in length
            *    21008    21107: gap of unknown length
            *    21108    31060: contig of 9953 bp in length
            *    31061    31160: gap of unknown length
            *    31161    49489: contig of 18329 bp in length
            *    49490    49589: gap of unknown length
            *    49590    99935: contig of 50346 bp in length
            *    99936   100035: gap of unknown length
            *   100036   198298: contig of 98263 bp in length.
FEATURES             Location/Qualifiers
     source          1. .198298
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-395L3"
     misc_feature    1. .1432
                     /note="assembly_name:Contig7"
     misc_feature    1533. .4137
                     /note="assembly_name:Contig9"
     misc_feature    4238. .7586
                     /note="assembly_name:Contig11"
     misc_feature    7687. .10403
                     /note="assembly_name:Contig12"
     misc_feature    10504. .21007
                     /note="assembly_name:Contig13"
     misc_feature    21108. .31060
                     /note="assembly_name:Contig14"
     misc_feature    31161. .49489
                     /note="assembly_name:Contig15"
     misc_feature    49590. .99935
                     /note="assembly_name:Contig16"
     misc_feature    100036. .198298
                     /note="assembly_name:Contig17"
BASE COUNT    55206 a  42154 c  42570 g  57542 t    826 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 198298;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2391 AATGGAAAACATTGTGTTTTTA 2412
              ||||||||||||||||||||||
Db     145746 AATGGAAAACATTGTGTTTTTA 145767


RESULT 46
AC122322/c
LOCUS       AC122322               203073 bp    DNA     linear   ROD 24-JAN-2003
DEFINITION  Mus musculus chromosome 6 clone RP23-321G20, complete sequence.
ACCESSION   AC122322
VERSION     AC122322.4  GI:27884993
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 203073)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 203073)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
```

```
  TITLE '   Direct Submission
  JOURNAL   Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 203073)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 203073)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Jan 24, 2003 this sequence version replaced gi:25167263.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information -----------------
            Center project name: M_BA0321G20
            ------------------------------------------------------.
FEATURES             Location/Qualifiers
     source          1. .203073
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="6"
                     /clone="RP23-321G20"
BASE COUNT    64340 a  39363 c  40150 g  59220 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 10;  Length 203073;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2372 AAATAAATAAAAACAAGCAAAT 2393
              |||||||||||||||||||||
Db     166976 AAATAAATAAAAACAAGCAAAT 166955


RESULT 47
AC132141
LOCUS       AC132141              205379 bp    DNA     linear   HTG 20-MAY-2003
DEFINITION  Mus musculus chromosome UNK clone RP23-55P19, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC132141
VERSION     AC132141.2  GI:30911164
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 205379)
  AUTHORS   Wilson,R.K.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 205379)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 205379)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAY-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On May 20, 2003 this sequence version replaced gi:22539405.

            ---------------- Genome Center ---------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information ------------------
            Center project name: M_BA0055P19
            ------------------------------------------------------
            ------------- Summary Statistics  -------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 203821 bases at least Q40
            Consensus quality: 204061 bases at least Q30
            Consensus quality: 204241 bases at least Q20
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   143221: contig of 143221 bp in length
            *   143222   143321: gap of unknown length
            *   143322   144526: contig of 1205 bp in length
            *   144527   144626: gap of unknown length
            *   144627   205379: contig of 60753 bp in length.
FEATURES             Location/Qualifiers
     source          1. .205379
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-55P19"
     misc_feature    1. .143221
                     /note="assembly_name:Contig10"
     misc_feature    143322. .144526
                     /note="assembly_name:Contig8"
     misc_feature    144627. .205379
                     /note="assembly_name:Contig9"
BASE COUNT    65292 a  39944 c  39561 g  60296 t    286 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 205379;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4240 TGACATTTTACTCCAAACTATT 4261
              |||||||||||||||||||||
Db       2614 TGACATTTTACTCCAAACTATT 2635


RESULT 48
AL773582
LOCUS       AL773582              206309 bp    DNA     linear   ROD 07-NOV-2002
DEFINITION  Mouse DNA sequence from clone RP23-174J23 on chromosome 2, complete
            sequence.
ACCESSION   AL773582
VERSION     AL773582.6  GI:23953985
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 206309)
  AUTHORS   Bates,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 13, 2002 this sequence version replaced gi:22798265.
            ---------- Genome Center
```

```
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-174J23 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .206309
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-174J23"
                     /clone_lib="RPCI-23"
BASE COUNT    65717 a  37351 c  38324 g  64917 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 10;  Length 206309;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1355 CTTTAAATGGAAAAACAAAAAA 1376
              ||||||||||||||||||||||
Db      92416 CTTTAAATGGAAAAACAAAAAA 92437


RESULT 49
AC122918/c
LOCUS       AC122918              212140 bp    DNA     linear   HTG 13-SEP-2002
DEFINITION  Mus musculus chromosome UNK clone RP23-29F15, WORKING DRAFT
            SEQUENCE, 6 unordered pieces.
ACCESSION   AC122918
VERSION     AC122918.3  GI:22748603
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 212140)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 212140)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 212140)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Sep 6, 2002 this sequence version replaced gi:22476048.

            ---------------- Genome Center ----------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information ------------------
            Center project name: M_BA0029F15
            ----------------------------------------------------
            ------------- Summary Statistics  -------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 209763 bases at least Q40
            Consensus quality: 210352 bases at least Q30
            Consensus quality: 210822 bases at least Q20
            Insert size: 192000; agarose-fp
            Insert size: 212858; sum-of-contigs
            Quality coverage: 13.40 in Q20 bases; agarose-fp
            Quality coverage: 11.34 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1408: contig of 1408 bp in length
            *     1409     1508: gap of unknown length
            *     1509     2809: contig of 1301 bp in length
            *     2810     2909: gap of unknown length
            *     2910    13613: contig of 10704 bp in length
            *    13614    13713: gap of unknown length
            *    13714    34255: contig of 20542 bp in length
            *    34256    34355: gap of unknown length
            *    34356    90086: contig of 55731 bp in length
            *    90087    90186: gap of unknown length
            *    90187   212140: contig of 121954 bp in length.
FEATURES             Location/Qualifiers
     source          1. .212140
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-29F15"
     misc_feature    1. .1408
                     /note="assembly_name:Contig18"
     misc_feature    1509. .2809
                     /note="assembly_name:Contig19"
     misc_feature    2910. .13613
                     /note="assembly_name:Contig20"
     misc_feature    13714. .34255
                     /note="assembly_name:Contig21"
     misc_feature    34356. .90086
                     /note="assembly_name:Contig22"
     misc_feature    90187. .212140
                     /note="assembly_name:Contig23"
BASE COUNT    62047 a  40079 c  41005 g  68508 t    501 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 212140;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2372 AAATAAATAAAAACAAGCAAAT 2393
              ||||||||||||||||||||||
Db       5676 AAATAAATAAAAACAAGCAAAT 5655
```

```
RESULT 50
AC131319
LOCUS       AC131319              225220 bp    DNA     linear   HTG 22-MAR-2003
DEFINITION  Mus musculus clone RP23-398N22, WORKING DRAFT SEQUENCE, 10
            unordered pieces.
ACCESSION   AC131319
VERSION     AC131319.3  GI:29150494
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 225220)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-398N22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 225220)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 225220)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 22, 2003 this sequence version replaced gi:28626750.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L26985
                Center clone name: 398_N_22
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 221437 bases at least Q40
                Consensus quality: 222996 bases at least Q30
                Consensus quality: 223710 bases at least Q20
                Insert size: 194000; agarose-fp
                Insert size: 224320; sum-of-contigs
                Quality coverage: 8.7 in Q20 bases; agarose-fp
                Quality coverage: 7.5 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2169: contig of 2169 bp in length
            *     2170     2269: gap of 100 bp
            *     2270     4247: contig of 1978 bp in length
            *     4248     4347: gap of 100 bp
            *     4348     8013: contig of 3666 bp in length
            *     8014     8113: gap of 100 bp
            *     8114    75957: contig of 67844 bp in length
            *    75958    76057: gap of 100 bp
            *    76058    88510: contig of 12453 bp in length
            *    88511    88610: gap of 100 bp
            *    88611   109038: contig of 20428 bp in length
            *   109039   109138: gap of 100 bp
            *   109139   130928: contig of 21790 bp in length
            *   130929   131028: gap of 100 bp
            *   131029   168658: contig of 37630 bp in length
            *   168659   168758: gap of 100 bp
            *   168759   223784: contig of 55026 bp in length
            *   223785   223884: gap of 100 bp
            *   223885   225220: contig of 1336 bp in length.
FEATURES             Location/Qualifiers
     source          1. .225220
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-398N22"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .2169
                     /note="assembly_fragment"
     misc_feature    2270. .4247
                     /note="assembly_fragment"
     misc_feature    4348. .8013
                     /note="assembly_fragment"
     misc_feature    8114. .75957
                     /note="assembly_fragment"
     misc_feature    76058. .88510
                     /note="assembly_fragment"
     misc_feature    88611. .109038
                     /note="assembly_fragment"
     misc_feature    109139. .130928
                     /note="assembly_fragment"
     misc_feature    131029. .168658
                     /note="assembly_fragment"
     misc_feature    168759. .223784
                     /note="assembly_fragment"
     misc_feature    223885. .225220
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    64151 a  49283 c  48888 g  61998 t    900 others
ORIGIN
```

```
  Query Match             0.2%;  Score 22;  DB 2;  Length 225220;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6419 TAATTAAACTTTTTTTATTCAA 6440
              ||||||||||||||||||||||
Db      49625 TAATTAAACTTTTTTTATTCAA 49646


RESULT 51
AC104177/c
LOCUS       AC104177               225439 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-13604, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC104177
VERSION     AC104177.7  GI:30521133
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 225439)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 225439)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-DEC-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 225439)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24942574.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GIQT
                Center clone name: CH230-13604
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 199660 bases at least Q40
                Consensus quality: 203425 bases at least Q30
                Consensus quality: 206374 bases at least Q20
                Estimated insert size: 211910; sum-of-contigs estimation
                Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   224212: contig of 224212 bp in length
            *    224213   224312: gap of unknown length
            *    224313   225439: contig of 1127 bp in length.
FEATURES             Location/Qualifiers
     source          1. .225439
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-13604"
     misc_feature    1. .1599
                     /note="wgs_contig"
     misc_feature    19070. .20182
                     /note="wgs_contig"
     misc_feature    38961. .42921
                     /note="wgs_contig"
     misc_feature    71930. .74110
                     /note="wgs_contig"
BASE COUNT    63911 a  38096 c  38132 g  68119 t  17181 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 225439;
```

```
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3928 TTTAAAAATATTTTTGAGGCTT 3949
              ||||||||||||||||||||||
Db      63288 TTTAAAAATATTTTTGAGGCTT 63267


RESULT 52
AC118928
LOCUS       AC118928              237418 bp    DNA     linear   HTG 25-MAY-2003
DEFINITION  Mus musculus clone RP23-416J10, *** SEQUENCING IN PROGRESS ***, 3
            unordered pieces.
ACCESSION   AC118928
VERSION     AC118928.4  GI:31074927
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 237418)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-416J10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 237418)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 237418)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2003 this sequence version replaced gi:28460870.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L20314
                Center clone name: 416_J_10
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    11003: contig of 11003 bp in length
            *    11004    11103: gap of 100 bp
            *    11104    44404: contig of 33301 bp in length
            *    44405    44504: gap of 100 bp
            *    44505   237418: contig of 192914 bp in length.
FEATURES             Location/Qualifiers
     source          1. .237418
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-416J10"
                     /clone_lib="RPCI-23 Female Mouse BAC"
BASE COUNT    67137 a  46243 c  47965 g  75485 t    588 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 237418;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4801 TATATAATATGTATAGAATTAT 4822
              ||||||||||||||||||||||
Db      40133 TATATAATATGTATAGAATTAT 40154


RESULT 53
AC098167/c
LOCUS       AC098167              242309 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-17K2, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC098167
VERSION     AC098167.5  GI:30521850
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 242309)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
```

```
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 242309)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 242309)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:23101378.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GHET
                Center clone name: CH230-17K2
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                  Consensus quality: 226446 bases at least Q40
                  Consensus quality: 229335 bases at least Q30
                  Consensus quality: 231093 bases at least Q20
                  Estimated insert size: 241670; sum-of-contigs estimation
                  Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   241010: contig of 241010 bp in length
            *    241011   241110: gap of unknown length
            *    241111   242309: contig of 1199 bp in length.
FEATURES              Location/Qualifiers
     source           1. .242309
                      /organism="Rattus norvegicus"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:10116"
                      /clone="CH230-17K2"
     misc_feature     1. .1864
                      /note="wgs_contig"
BASE COUNT    64712 a  47608 c  49926 g  70307 t   9756 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 242309;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2371 TAAATAAATAAAAACAAGCAAA 2392
              ||||||||||||||||||||||
Db     170128 TAAATAAATAAAAACAAGCAAA 170107


RESULT 54
AC105871/c
LOCUS       AC105871               244698 bp    DNA     linear   HTG 08-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-150F21, *** SEQUENCING IN PROGRESS
            ***.
ACCESSION   AC105871
VERSION     AC105871.3  GI:22857136
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 244698)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
```

```
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 244698)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 244698)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 14, 2002 this sequence version replaced gi:21743866.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GNYK
                Center clone name: CH230-150F21
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 206902 bases at least Q40
                Consensus quality: 209753 bases at least Q30
                Consensus quality: 211564 bases at least Q20
                Estimated insert size: 223792; sum-of-contigs estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   244698: contig of 244698 bp in length.
FEATURES            Location/Qualifiers
     source         1. .244698
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-150F21"
     misc_feature   1. .1123
                    /note="wgs_contig"
     misc_feature   107508. .109511
                    /note="wgs_contig"
     misc_feature   151221. .152377
                    /note="wgs_contig"
BASE COUNT    61125 a  39764 c  42502 g  69289 t  32018 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 244698;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2371 TAAATAAATAAAAACAAGCAAA 2392
              ||||||||||||||||||||||
Db      90722 TAAATAAATAAAAACAAGCAAA 90701


RESULT 55
AE014819
LOCUS       AE014819              251448 bp    DNA     linear   INV 07-OCT-2002
DEFINITION  Plasmodium falciparum 3D7 chromosome 14 section 4 of 13 of the
            complete sequence.
ACCESSION   AE014819 AE014187
VERSION     AE014819.1  GI:23497243
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 251448)
  AUTHORS   Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,
            Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
            James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
            Kyes,S., Chan,M.-S., Nene,V., Shallom,S.J., Suh,B., Peterson,J.,
            Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
            Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H.,
            Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
            Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
            Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
            Barrell,B.
  TITLE     Genome sequence of the human malaria parasite Plasmodium falciparum
  JOURNAL   Nature 419 (6906), 498-511 (2002)
   PUBMED   12368864
REFERENCE   2  (bases 1 to 251448)
  AUTHORS   Gardner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES            Location/Qualifiers
     source         1. .251448
                    /organism="Plasmodium falciparum 3D7"
                    /mol_type="genomic DNA"
                    /isolate="3D7"
                    /db_xref="taxon:36329"
                    /chromosome="14"
     repeat_region  482. .1146
```

```
                /rpt_type=tandem
repeat_region   1144. .1188
                /rpt_type=tandem
repeat_region   1559. .1591
                /rpt_type=tandem
repeat_region   2017. .2049
                /rpt_type=tandem
repeat_region   2105. .2146
                /rpt_type=tandem
gene            complement(<2683. .>3858)
                /locus_tag="PF14_0178"
mRNA            complement(join(<2683. .2946,3183. .3223,3318. .3399,
                3532. .3625,3728. .>3858))
                /locus_tag="PF14_0178"
CDS             complement(join(2683. .2946,3183. .3223,3318. .3399,
                3532. .3625,3728. .3858))
                /locus_tag="PF14_0178"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAN36790.1"
                /db_xref="GI:23497244"
                /translation="MMQQLNLKEGDIVRVTSVSLPKGTFVKLKPCSKDFMELSNHRAV
                LETALRNYATLTIGDNIVIHYLGKTYEIKIVDLKPAFACTIIETDVEVEFEEPIDHSE
                PVQYVTEIVPEEENKFKGKGQRTDGKSCHKALKKEIIKPKVVENLEPWKDKLAQGIRV
                IRMINIYIYIYIYIFIYLFIYVCMYVFILFLFFRRNVLNMKIC"
repeat_region   2772. .2801
                /rpt_type=tandem
repeat_region   3245. .3290
                /rpt_type=tandem
repeat_region   4265. .4335
                /rpt_type=tandem
repeat_region   4614. .4676
                /rpt_type=tandem
repeat_region   5091. .5115
                /rpt_type=tandem
repeat_region   5573. .5617
                /rpt_type=tandem
repeat_region   5797. .5837
                /rpt_type=tandem
gene            complement(<5884. .>16677)
                /locus_tag="PF14_0179"
mRNA            complement(<5884. .>16677)
                /locus_tag="PF14_0179"
CDS             complement(5884. .16677)
                /locus_tag="PF14_0179"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAN36791.1"
                /db_xref="GI:23497245"
                /translation="MKLSVLYLVVSLYFFDRKIVCNKKNGKSKSNNHNYFNNTTKHID
                KQVIYNLPENDNLFHVLKYLIHGENKNILYDVDESLYTKIENKEYITKEDLLRTLALI
                EKKNVDCNDNIKKVKIAKLLIDINDLLRTNYIILDDYKEINKNDLTKNGFLLKLQNSL
                NKRDMVNSKRVNQDIILRNQYLEKNYKLCSISSDMLIISTPRNLNISDTSIKEIKHPS
                SINIMNEIFCDNVISSNVSNISPLQLNKSIGLYKPLLDLDIKSRIDEYRKKIITYYQD
                SMNEIFYLIRDEFNEIKKKFYLYIHPSKMPEAKPNNDIFKKEDMEGNEIITYVSNKIE
                NKENIKNKEEYILKSNNNNVIKISSKRTKKVINIPIKKKKLSNDIKENVNSSTKSDNN
                NIISNETVDACTKIESENGNKKKDDNVKNINRIYHIFLKKKNNEKKCDISKQTLYNNS
                FKENDVIKTFTHNDNILKKIIIIKDVSREKPTIKEIIHQQPLPVENSELYNISEHIQN
                KCNCCFYSIPLIHIYHPNGSMHCLSCSQYEKNVNRKCPIYEKYITFNYNDTCLLCVPS
                SYITNNCFCNLIHSKGLGEKTNDVPKKLNMMPHFSSIIDVYGNKNIDGYNNFENDNNL
                LVYSNNDEPILKKNHKERPNKKCSNRYKLKCPVGNGSSDMFSGQSISIFPIIGDDINR
                GHYENCPYYDEEDVMCYNTITNKKININNEDDNYEKILYDKDKIVYLNSSSFDTINIK
                IEGEDLNMNDRNEENDLLDKESIEYINEINEKEYYDSIEACKDYRYHTFEESIKNVNN
                RECYKNVYYSTVHKKEDNIDKKCSHDIESRNKRVNEYEDECENDNNNNNNNDDDDGE
                ERNYFLFDLRFIYKTGFFTKRKTIINSYLKTNKIDITYNKEKILYLFKNLFEKTFKDI
                NYQIKESLVHMFPIVQNSEIQGFDIDFGKVKLHYATSINMFKYSVFDLINMKIYFISP
                DKTYILLEDIIREIREEKENSNNDGSRRKNKIINKYLHEMDITNIEWSTENSKTNTTN
                SESKTTTTRENTIVIKEKIGTHDEKKKKCDNTKNNRNNNDDNDDNDNDNNNDNNND
                NNNNNNNNNNNNNNNNNNNNNNNSNSNSCDNNNKRDNDSFEHVDIQKNDTITTNIDCNI
                IHNNTIDSSNKNRIKYNLNILKFLLKYKDLLSKEDINMISDKNNLVTINKNDKTNIGD
                EIIEITIRKNEKRNIENNCTYTNNSIINNIDNNISESININGGEGINAVIETVDKYLD
                KYFSVHNMPIYYMENMKISKNIIYVTSDVNSDLDMLKMSILGICNITNKSIDNFNFFL
                VRNEAELYDPNKCQKIPEHIKNVQDLYNYSKMLDMKIIIAPVYAEDIDIIPKSFDYLF
                KLYKDGDDEKGVTTDSRDNDCQQKGNVDTQERNKTENDTTRITLSNIKHNNIIKNNNK
                NNNNNNNNSHNNNHSYYNHELSLFNDAESNKPNDNTTTSNDKNYYFFNYDNKKSLILN
                DPENDENIEYLSEEISIYEDEHIYDEGNLYNSFVIYIDNPFKRINIENAPILKGYTLN
                LFINDIMNKLIHIPSTFINTQSYKKGMCVMSICYCFNYYSLNIYTNNIILLNFDKDVT
                LYEVIKEFKSRENEYLLIKNKMNEECQDVDNLSNIKLPHTIDLIDSKYIHLEFPLLYI
                DEHSNFFQDKITLMNVPENLTVYELCKSMKNILNESLISFNLSTINDIKIYTIRGRQW
                EMVLDNLLVFGLKINYEDIFTILIDNNLLTKNYYELLSNIIIDFRDNNMYIKKLDIYI
                DYIFQPYTIYNIPIYITPNNLKYILLQYTNTFLSNSLLDEFFFTYNKRHTTINVVPEV
                RNWEHVEDSDVYIYEENDINDKKKEQKRNISFIYVLNMLSTFYKIKFNSSKLLESVMG
                EIHQLCGKRFNENLWLDIKKSTKNEKDIEIYLENSSKKVLIKNVPLDLLVWKLINIIM
                RGSFNIPIEDKEKLYNLFMLVPKNKDKIKDKNVDYYFLSTPGYTVEEMLRVFEKKKKP
                KKKGKEKNSANDNNNNNNNDNNNDNYNDNYNNDNNINSNNNYGDMKQTYYKILLIKAY
                PQHLHGIQNNERFNEYLTFNLEALPSVIDVKNPTTCLSFYVNYKNENKMTHIINVPEN
                ISTDMLLKTILSDNFCRWKKLPNEEKMKFSLHLNKKEITNIKDYINNGDIAQLKTFIL
                INNKKRRVQEKYIDLDNTTELGSNVNNISKEELKEIQLYDNVLREILNKNYIDYNNIS
                LSGYTITVYVFDGNKYEPLTIYNVPLNITNKDIINFLLIKSNHINITKLIDEFLLLNE
                KLYNTYIVKNKNKNMLEEQNVLFIKKLNEEEEVFQLNESKFYLVHNPLFNPLDNLLQN
                IVHKKYDFKSEKIDIHSVGEKFEILTLNILINNDINNMHNNNVNNNYVNNNNINNNNI
                NNNNINNNNINNNNINNNNFSFGKKKFVITVLNIPYNMYITEFLRYSIGYQRKWIYKF
                VLFYIIKGSEKYILNDGNDIIRHGKTVSEIFQICKSGDLCTLHIDINFNEADNNLEND
                IFNEEINAKEMKNVDVDSHADVDVDLDANVYLDANADVEGNINKFVNSLHNDNENRFI
                VLKELSKKIYTDDYENKKIDMSIICFEYNAGLYERNLFGSSRICNIPKNYTVGDVLRY
                VKNKIIEDTNISMLNDLQLKIIHNDNYIDIPKDMNIEYAINYLFIYTPFIVYPSDDIL
                KYNNIIDLILYNTIIDIKENSFVKKDIPIYVKKGNIFKNVSLKYIPINIKEDELLTYL
                LHYLTDNHEVVSFIRDITCICVYPLCDSIYIFHEKKQLSLIPSISYHIQLKNVYYIST
                VETYENFYNILGHMELNLNKFLHIQNDGNNYKKKEAYNNIYYMNNVDELIHIDVNIIF
                PSIQKIIRVKNLNVNIKMEELLYKIFGCITNKIYKWGELFRNFVGKKEKINIVNNDIK
                NNSNNINSINNMNSINNMNNINNMNSINNINNINSINNINNCDNIDRIDNKVLKDYLC
                EEQNIITFIFCSHDNSIHDYLTSRMIYLPAIINYQTNYITFYSSINGFANNETTFHGF
                PNYLSPETMLTMLQYNFFHVIGKNYLKLYGSLYDKGTDEDIIYNTNDIRKSENKDIYN
                KNLIAMRGAIKFILRIKNKGKKKSLDTIQNLINNELNIACENVENDVDMEECNDFLSS
                LNNSSIFWKNGILGFMANSIVTRDNMNNSLILPYVDFEINEKMFLQNILYHINISPYL
                YKLFELTHIDKKCNTYKKKLKNSVAHIQCLLSYGFSIYFDLHYDSTFDKYKGFTLGDN
                SFDITQIHSFQNLHEQNFVDLILYNHDNTFITIKNVYRDMTVATFLNDLSKTRCHILG
                LTYNQISTYYKLTYEIYNETHDISPNLSIADVHNIVISFSKSSILLKIRKIDITNDIS
                HNNNNNNNNDSDFIINSDNIDYINALPDLYLSPYPPVIDLSQNNFQVNIYMHSKFIPN
                DYSLRKDIVPHITINNVPSNVFILSLKELLKTIFIQHMEKSQTTTFQLNLYDFEVEFF
                ILTFNRFTSKFHKMNSNVLSNLTISNFYSTYIHGGLFLNIDKIKIDIPNLLNDFIDHF
                SKVVIDFSHHNIVKQ"
repeat_region   6427. .6452
                /rpt_type=tandem
repeat_region   7723. .7842
                /rpt_type=tandem
repeat_region   7723. .7837
                /rpt_type=tandem
repeat_region   9355. .9481
                /rpt_type=tandem
repeat_region   9355. .9441
                /rpt_type=tandem
repeat_region   10507. .10600
                /rpt_type=tandem
repeat_region   13335. .13476
                /rpt_type=tandem
repeat_region   14128. .14153
                /rpt_type=tandem
repeat_region   16847. .16882
                /rpt_type=tandem
repeat_region   17437. .17467
                /rpt_type=tandem
repeat_region   17854. .17884
                /rpt_type=tandem
repeat_region   18281. .18332
                /rpt_type=tandem
repeat_region   18538. .18585
                /rpt_type=tandem
gene            <18730. .>19104
                /locus_tag="PF14_0180"
mRNA            <18730. .>19104
                /locus_tag="PF14_0180"
CDS             18730. .19104
                /locus_tag="PF14_0180"
                /codon_start=1
                /product="hypothetical protein"
```

```
                     /protein_id="AAN36792.1"
                     /db_xref="GI:23497246"
                     /translation="MSETNKQNDNKKNKITRESLLDDQFSENPSSKFVLNKEEKKHEE
                     INPNISQHKVNKKKPLIDRSLIGKFSVEKYKSFLSKIEKSNEDIKNKDAESVRIDKDI
                     LNSECDNNNNNNESCVYMGNLF"
     repeat_region   19647. .19685
                     /rpt_type=tandem
     repeat_region   19815. .19846
                     /rpt_type=tandem
     gene            <19942. .>20352
                     /locus_tag="PF14_0181"
     mRNA            <19942. .>20352
                     /locus_tag="PF14_0181"
     CDS             19942. .20352
                     /locus_tag="PF14_0181"
                     /codon_start=1
                     /product="calmodulin, putative"
                     /protein_id="AAN36793.1"

  Query Match             0.2%;  Score 22;  DB 3;  Length 251448;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6702 AAAAAAAACAAAGATTTAAAAA 6723
             ||||||||||||||||||||||
Db      58425 AAAAAAAACAAAGATTTAAAAA 58446


RESULT 56
AC096611
LOCUS       AC096611               252749 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-54A16, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC096611
VERSION     AC096611.9  GI:30522099
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 252749)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 252749)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 252749)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:23321952.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GGBT
                Center clone name: CH230-54A16
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 246639 bases at least Q40
                Consensus quality: 248286 bases at least Q30
                Consensus quality: 249278 bases at least Q20
                Estimated insert size: 258972; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   250721: contig of 250721 bp in length
```

```
            *   250722   250821: gap of unknown length
            *   250822   252749: contig of 1928 bp in length.
FEATURES             Location/Qualifiers
     source          1. .252749
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-54A16"
     misc_feature    6561. .7174
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH344696"
     misc_feature    complement(247773. .248640)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH344695"
BASE COUNT     83784 a  50332 c  47503 g  68128 t   3002 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 252749;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2371 TAAATAAATAAAAACAAGCAAA 2392
              ||||||||||||||||||||||
Db      16009 TAAATAAATAAAAACAAGCAAA 16030


RESULT 57
AC109408
LOCUS       AC109408              267093 bp    DNA     linear   HTG 09-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-270I12, *** SEQUENCING IN PROGRESS
            ***, 8 unordered pieces.
ACCESSION   AC109408
VERSION     AC109408.4  GI:23101181
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 267093)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 267093)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 267093)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 18, 2002 this sequence version replaced gi:21737978.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GQBF
                Center clone name: CH230-270I12
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 205376 bases at least Q40
                Consensus quality: 210255 bases at least Q30
                Consensus quality: 213384 bases at least Q20
                Estimated insert size: 233774; sum-of-contigs estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
```

```
            *        1    10796: contig of 10796 bp in length
            *    10797    10896: gap of unknown length
            *    10897   256231: contig of 245335 bp in length
            *   256232   256331: gap of unknown length
            *   256332   257997: contig of 1666 bp in length
            *   257998   258097: gap of unknown length
            *   258098   259380: contig of 1283 bp in length
            *   259381   259480: gap of unknown length
            *   259481   261177: contig of 1697 bp in length
            *   261178   261277: gap of unknown length
            *   261278   262887: contig of 1610 bp in length
            *   262888   262987: gap of unknown length
            *   262988   265407: contig of 2420 bp in length
            *   265408   265507: gap of unknown length
            *   265508   267093: contig of 1586 bp in length.
FEATURES             Location/Qualifiers
     source          1. .267093
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-270I12"
     misc_feature    1. .1854
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    9659. .10796
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    9807. .10667
                     /note="clone_boundary
                     clone_end:Sp6
                     site:MboI
                     end_sequence:RXAEV54TV"
     misc_feature    10897. .13096
                     /note="wgs_contig"
     misc_feature    201191. .202025
                     /note="clone_boundary
                     clone_end:T7
                     site:MboI
                     end_sequence:RXAEV54TJ"
BASE COUNT    68130 a  42943 c  41795 g  62177 t  52048 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 2;  Length 267093;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2371 TAAATAAATAAAAACAAGCAAA 2392
             ||||||||||||||||||||||
Db      90784 TAAATAAATAAAAACAAGCAAA 90805


RESULT 58
BV028873
LOCUS       BV028873                845 bp    DNA     linear   STS 30-MAY-2003
DEFINITION  S212P6333FF12.T0 CZECHII/Ei Mus musculus STS genomic, sequence
            tagged site.
ACCESSION   BV028873
VERSION     BV028873.1  GI:31112768
KEYWORDS    STS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 845)
  AUTHORS   Wade,C.M., Kulbokas,E.J. III, Kirby,A.W., Zody,M.C., Mullikin,J.C.,
            Lander,E.S., Lindblad-Toh,K. and Daly,M.J.
  TITLE     The mosaic structure of variation in the laboratory mouse genome
  JOURNAL   Nature 420 (6915), 574-578 (2002)
  MEDLINE   22354684
   PUBMED   12466852
COMMENT
            Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kersli@genome.wi.mit.edu
            Primer A: None
            Primer B: None
            STS size: 845
            Protocol:
            WGS-discovery: Paired-end low-coverage whole genome shotgun reads
            were generated from 129S1/SvImJ, C3H/HeJ, and BALB/cByJ. The WGS
            reads were placed uniquely on the MGSCv3 C57BL/6J assembly and SNP
            detection was carried out by SSAHA-SNP. 225,000 reads were
            annotated
            as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J
            and the strain from which the particular read came. The validation
            rate for these SNPs was estimated at approximately 98%.
FEATURES             Location/Qualifiers
     source          1. .845
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="CZECHII/Ei"
                     /db_xref="taxon:10090"
                     /map="+ 11 22-800 47879139-47879919"
                     /clone_lib="CZECHII/Ei"
     STS             <1. .>845
BASE COUNT      227 a    184 c    165 g    269 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 11;  Length 845;
  Best Local Similarity   100.0%;  Pred. No. 65;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7940 CTTGTTTTTCAACCCATCTTT 7960
              |||||||||||||||||||||
Db        696 CTTGTTTTTCAACCCATCTTT 716


RESULT 59
AX598771/c
LOCUS       AX598771                3628 bp    DNA     linear   PAT 14-FEB-2003
DEFINITION  Sequence 111 from Patent WO02077272.
ACCESSION   AX598771
VERSION     AX598771.1  GI:28398909
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Berlin,K., Braun,A., Distler,J., Guetig,D., Howe,A., Mueller,J.,
            Olek,A., Piepenbrock,C., Adorjan,P., Grabs,G., Lesche,R., Leu,E.,
            Lewin,A., Lipscher,E., Maier,S., Model,F., Mueller,V., Otto,T.,
            Pelet,C. and Ziebarth,H.
  TITLE     Methods and nucleic acids for the analysis of hematopoietic cell
            proliferative disorders
  JOURNAL   Patent: WO 02077272-A 111 03-OCT-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .3628
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT      734 a    246 c   1028 g   1620 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 6;  Length 3628;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2367 CTACTAAATAAATAAAAACAA 2387
              |||||||||||||||||||||
Db       1013 CTACTAAATAAATAAAAACAA 993
```

```
RESULT 60
AX598917/c
LOCUS       AX598917               3628 bp    DNA     linear   PAT 14-FEB-2003
DEFINITION  Sequence 257 from Patent WO02077272.
ACCESSION   AX598917
VERSION     AX598917.1  GI:28399057
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Berlin,K., Braun,A., Distler,J., Guetig,D., Howe,A., Mueller,J.,
            Olek,A., Piepenbrock,C., Adorjan,P., Grabs,G., Lesche,R., Leu,E.,
            Lewin,A., Lipscher,E., Maier,S., Model,F., Mueller,V., Otto,T.,
            Pelet,C. and Ziebarth,H.
  TITLE     Methods and nucleic acids for the analysis of hematopoietic cell
            proliferative disorders
  JOURNAL   Patent: WO 02077272-A 257 03-OCT-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .3628
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT      734 a      0 c   1028 g   1866 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 6;  Length 3628;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2367 CTACTAAATAAATAAAAACAA 2387
             |||||||||||||||||||||
Db       1013 CTACTAAATAAATAAAAACAA 993


RESULT 61
AX345495/c
LOCUS       AX345495               6127 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 566 from Patent WO0200928.
ACCESSION   AX345495
VERSION     AX345495.1  GI:18493381
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 566 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .6127
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT     1578 a     67 c   1387 g   3095 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 6;  Length 6127;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2367 CTACTAAATAAATAAAAACAA 2387
             |||||||||||||||||||||
Db       2821 CTACTAAATAAATAAAAACAA 2801


RESULT 62
ECO518835
LOCUS       ECO518835              10910 bp    DNA     linear   BCT 27-NOV-2002
DEFINITION  Escherichia coli pMBSF1 plasmid.
ACCESSION   AJ518835
VERSION     AJ518835.1  GI:25815144
KEYWORDS    floR gene; florfenicol/chloramphenicol resistance protein; ORF4;
            ORF6; ORF8; ORF9; strA gene; strB gene; streptomycin resistance
            protein A; streptomycin resistance protein B; tnp gene;
            transposase.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Blickwede,M. and Schwarz,S.
  TITLE     Molecular analysis of the florfenicol resistance gene region of
            plasmid pMBSF1 from porcine E. coli
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 10910)
  AUTHORS   Schwarz,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-2002) Schwarz S., Molecular Microbiology &
            Diagnostics, Institute for Animal Science (FAL), Hoeltystr. 10,
            Neustadt-Mariensee, D-31535, GERMANY
FEATURES             Location/Qualifiers
     source          1. .10910
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:562"
                     /lab_host="Sus scrofa"
                     /plasmid="pMBSF1"
                     /country="Germany"
     gene            723. .1526
                     /gene="strA"
     CDS             723. .1526
                     /gene="strA"
                     /codon_start=1
                     /transl_table=11
                     /product="streptomycin resistance protein A"
                     /protein_id="CAD57666.1"
                     /db_xref="GI:25815145"
                     /translation="MNRTNIFFGESHSDWLPVRGGESGDFVFRRGDGHAFAKIAPASR
                     RGELAGERDRLIWLKGRGVACPEVINWQEEQEGACLVITAIPGVPAADLSGADLLKAW
                     PSMGQQLGAVHSLSVDQCPFERRLSRMFGRAVDVVSRNAVNPDFLPDEDKSTPQLDLL
                     ARVERELPVRLDQERTDMVVCHGDPCMPNFMVDPKTLQCTGLIDLGRLGTADRYADLA
                     LMIANAEENWAAPDEAERAFAVLFNVLGIEAPDRERLAFYLRLDPLTWG"
     gene            1526. .2362
                     /gene="strB"
     CDS             1526. .2362
                     /gene="strB"
                     /codon_start=1
                     /transl_table=11
                     /product="streptomycin resistance protein B"
                     /protein_id="CAD57667.1"
                     /db_xref="GI:25815146"
                     /translation="MFMPPVFPAHWHVSQPVLIADTFSSLVWKVSLPDGTPAIVKGLK
                     PIEDIADELRGADYLVWRNGRGAVRLLGRENNLMLLEYAGERMLSHIVAEHGDYQATE
                     IAAELMAKLYAASEEPLPSALLPIRDRFAALFQRARDDQNAGCQTDYVHAAIIADQMM
                     SNASELRGLHGDLHHENIMFSSRGWLVIDPVGLVGEVGFGAANMFYDPADRDDLCLDP
                     RRIAQMADAFSRALDVDPRRLLDQAYAYGCLSAAWNADGEEEQRDLAIAAAIKQVRQT
                     SY"
     gene            2486. .3409
                     /gene="tnp"
     CDS             2486. .3409
                     /gene="tnp"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase"
                     /protein_id="CAD57668.1"
                     /db_xref="GI:25815147"
                     /translation="MLFLDGVYVEQSHGSARFRWVKAPTSPELTQLTHTIAHRVGRYL
                     ERQGLLERDVENSYLASDAVDDDPMTPLLGHSITYRIAVGSQAGRKVFTLQTLPTSGD
                     PFGDGIGKVAGSSLHAGVAARADERKKLERLCRYISRPAVSEKRLSLTRGGNVRYQLK
```

```
                TPYRDGTTHVIFEPLDFIARLAALVPKPRVNLTRFHGVFAPNSRHRALVTPAKRGRGN
                KVRVADEPATPAQRRASMTWAQRLKRVFNIDIETCSGCGGAMKVIACIEDPIVIKQIL
                DHLKHKAETSGTRALPESRAPPAELLLGLFD"
     CDS        3440. .4324
                /note="ORF4"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="CAD57669.1"
                /db_xref="GI:25815148"
                /translation="MLRSQRRGRDPPCWLSEKEPLVGKRRVNFQRCWLPVSRIGLHRR
                GVERVNCGPKLLDLGEKGVYSSYTCRVPGGDYEGYVDAHVRRLEALRRAGIVERIDAD
                QWRIPDDLVSRAAAHDAGRDSQASVRVLSPVDLNKQIGSDGATWLDRRLIHGETADLA
                PTGFGQQVREAMDQRREHHIEQGDATRSRDSRVFYRRNLLAILREREVAGVGSDMALS
                KGLPFRAATDGESVSGKFTGTVHLSSGKFAVVEKSHEFTLVPWRPIIDRQLGREVMGI
                VQGGSVSWQLGRQRGLER"
     gene       4535. .5749
                /gene="floR"
     CDS        4535. .5749
                /gene="floR"
                /codon_start=1
                /transl_table=11
                /product="florfenicol/chloramphenicol resistance protein"
                /protein_id="CAD57670.1"
                /db_xref="GI:25815149"
                /translation="MTTTRPAWAYTLPAALLLMAPFDILASLAMDIYLPVVPAMPGIL
                NTTPAMIQLTLSLYMVMLGVGQVIFGPLSDRIGRRPILLAGATAFVIASLGAAWSSTA
                PAFVAFRLLQAVGASAMLVATFATVRDVYANRPEGVVIYGLFSSILAFVPALGPIAGA
                LIGEFLGWQAIFITLAILAMLALLNAGFRWHETRPLDQVKTRRSVLPIFASPAFWVYT
                VGFSAGMGTFFVFFSTAPRVLIGQAEYSEIGFSFAFATVALVMIVTTRFAKSFVARWG
                IAGCVARGMALLVCGAVLLGIGELYGSPSFLTFILPMWVVAVGIVFTVSVTANGALAE
                FDDIAGSAVAFYFCVQSLIVSIVGTLAVALLNGDTAWPVICYATAMAVLVSLGLVLLR
                LRGAATEKSPVV"
     CDS        5777. .6082
                /note="ORF6"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="CAD57671.1"
                /db_xref="GI:25815150"
                /translation="MRRTNHRNLVNVGILSGRIPLISLVQFIAVAEHLNFRHAAKALG
                ISQSSVSARVKALEDNLGVLLFERHARGVRLTDAGRHFMERVTAGVDQLDHAVKTAE"
     gene       7038. .8858
                /gene="tnp"
     CDS        7038. .8858
                /gene="tnp"
                /codon_start=1
                /transl_table=11
                /product="transposase"
                /protein_id="CAD57672.1"
                /db_xref="GI:25815151"
                /translation="MAIRRRRRMFILFAERKVGEQHGPAAQGVLAAVQTLREMNADNL
                RKVPADAPTAFIKPRWKPLVITPEGLDRKFYEICALSELKNALRSGDIWVKGSRQFRD
                FDDYLLPAEKFAALKREQALPLAINPNSDQYLEERLQLLDEQLATVTRLAKDNELPDA
                ILTESGLKITPLDAAVPDRAQALIDQTSQLLPRIKITELLMDVDDWTGFSRHFTHLKD
                GAEAKDRTLLLSAILGDAINLGLTKMAESSPGLTYAKLSWLQAWHIRDETYSAALAEL
                VNHQYRHAFAAHWGDGTTSSSDGQRFRAGGRGESTGHVNPKYGSEPGRLFYTHISDQY
                APFSTRVVNVGVRDSTYVLDGLLYHESDLRIEEHYTDTAGFTDHVFALMHLLGFRFAP
                RIRDLGETKLYVPQGVQAYPTLRPLIGGTLNIKHVRARWDDILRLASSIKQGTVTASL
                MLRKLGSYPRQNGLAVALRELGRIERTLFILDWLQSVELRRRVHAGLNKGEARNSLAR
                AVFFNRLGEIRDRSFEQQRYRASGLNLVTAAIVLWNTVYLERATQGLVEAGKPVDGEL
                LQFLSPLGWEHINLTGDYVWRQSRRLEDGKFRPLRMPGKP"
     CDS        complement(9398. .10093)
                /note="ORF8"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="CAD57673.1"
                /db_xref="GI:25815152"
                /translation="MKVAVINFSGNVGKSTVAQHLLAPRMPNAEFVAVESINADEGTG
                EVVRGKQFGQLQEQLLMVDEAVIDIGASNVEDFVKLMQQYRGSHEDFDYFVVPAVKES
                KQIKDTIATINALSKMGVPAKKIRVVFNKLEADETVEEAFYPLFAYHDDSKAFTLKPK
                AAIQFSELYHKLRTYRVTIEQLLTDETDYKAKLREAKDPEEKAQAAAMISMRRLAYSA
                KENLDAVYAAISK"
     CDS        complement(10090. .10521)
                /note="ORF9"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="CAD57674.1"
                /db_xref="GI:25815153"
                /translation="MQVSETDALKDRIENIRPRMTLAAKLRELMPEIDRQVRAGVQHD
                DIVETLNANGFDVNLNTFRSYLYRYRKKARASEGQPEPAQATPKADGNSPVVSEGPGD
                VSPDFDADSLADALDARKRDELGDKYLGLKRPIFSKRSDNK"
BASE COUNT      2192 a    3204 c    3230 g    2284 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 1;  Length 10910;
  Best Local Similarity   100.0%;  Pred. No. 74;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2710 CGCCCGTGCCCTCATCGGCGT 2730
              |||||||||||||||||||||
Db       9961 CGCCCGTGCCCTCATCGGCGT 9981


RESULT 63
AC110783/c
LOCUS       AC110783               23481 bp    DNA     linear   PRI 16-APR-2002
DEFINITION  Homo sapiens BAC clone RP11-390P9 from 4, complete sequence.
ACCESSION   AC110783
VERSION     AC110783.5  GI:19909445
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 23481)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 23481)
  AUTHORS   Harris,A. and Haakenson,W.
  TITLE     The sequence of Homo sapiens BAC clone RP11-390P9
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 23481)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 23481)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 23481)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 23481)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Apr 3, 2002 this sequence version replaced gi:19551190.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
```

```
              ---------- Summary Statistics
                  Center project name: H_NH0390P09
              ----------.

              NOTICE:  This sequence may not represent the entire insert of this
              clone.  It may be shorter because we only sequence overlapping
              clone sections once, or longer because we provide a small overlap
              between neighboring data submissions.

              This sequence was finished as follows unless otherwise noted:
              all regions were double stranded, sequenced with an alternate
              chemistry, or covered by high quality data (i.e., phred quality >=
              30); an attempt was made to resolve all sequencing problems, such
              as compressions and repeats; all regions were covered by sequence
              from more than one subclone; and the assembly was confirmed by
              restriction digest.

              MAPPING INFORMATION:
              Mapping information for this clone was provided by Dr. John D.
              McPherson, Department of Genetics, Washington University, St. Louis
              MO. For additional information about the map position of this
              sequence, see http://genome.wustl.edu/gsc

              SOURCE INFORMATION:
              The RPCI-11 human BAC library was made from the blood of one male
              donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
              Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
              approach for construction of bacterial artificial chromosome
              libraries. Genomics 51:1-8. The clone may be obtained either from
              Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
              and coworkers at http://www.chori.org
              VECTOR:  pBACe3.6

              NEIGHBORING SEQUENCE INFORMATION:
              The clone sequenced to the left is RP11-1J11, 2000 bp overlap; the
              clone sequenced to the right is RP11-292F20, 2000 bp overlap.
              Actual start of this clone is at base position 35778 of RP11-1J11;
              actual end is at base position 46245 of RP11-292F20.
FEATURES              Location/Qualifiers
     source           1. .23481
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="4"
                      /map="4"
                      /clone="RP11-390P9"
                      /clone_lib="RPCI-11"
     repeat_region    2443. .2799
                      /rpt_family="MaLR"
     repeat_region    4130. .4267
                      /rpt_family="L1"
     repeat_region    4290. .4386
                      /rpt_family="MaLR"
     repeat_region    5307. .5352
                      /rpt_family="AT_rich"
     misc_feature     5674. .5814
                      /note="similar to EST BM429530 (NID:g18451252)"
     misc_feature     5847. .6137
                      /note="similar to EST BM429530 (NID:g18451252)"
     repeat_region    6285. .6691
                      /rpt_family="MaLR"
     repeat_region    6731. .6811
                      /rpt_family="MaLR"
     repeat_region    6813. .6845
                      /rpt_family="(TG)n"
     repeat_region    7270. .7381
                      /rpt_family="MIR"
     repeat_region    7426. .7570
                      /rpt_family="MIR"
     repeat_region    8315. .8568
                      /rpt_family="MIR"
     repeat_region    11318. .12682
                      /rpt_family="L1"
     repeat_region    13382. .13672
                      /rpt_family="Alu"
     repeat_region    15934. .16173
                      /rpt_family="L1"
     repeat_region    16485. .16779
                      /rpt_family="Alu"
     repeat_region    16943. .17068
                      /rpt_family="L2"
     repeat_region    17197. .18240
                      /rpt_family="MER2_type"
     repeat_region    18311. .18336
                      /rpt_family="AT_rich"
     repeat_region    18369. .18738
                      /rpt_family="L1"
     repeat_region    18739. .18760
                      /rpt_family="Alu"
     repeat_region    18761. .18786
                      /rpt_family="(TTAA)n"
     repeat_region    18787. .19073
                      /rpt_family="Alu"
     repeat_region    19074. .19161
                      /rpt_family="L1"
     repeat_region    19254. .19357
                      /rpt_family="(TATATG)n"
     repeat_region    19441. .19521
                      /rpt_family="CT-rich"
     misc_feature     19902
                      /note="match to EST BG433965 (NID:g13340471)"
     misc_feature     19967. .20449
                      /note="similar to Mus musculus EST AI931633 (NID:g5668118)
                      ul71e12.y1"
     misc_feature     19993. .20449
                      /note="match to EST BG433965 (NID:g13340471)"
     misc_feature     20206. .20424
                      /note="similar to Mus musculus EST AW044772 (NID:g5905241)
                      um11a07.y1"
     repeat_region    20877. .21187
                      /rpt_family="Alu"
     repeat_region    21224. .21532
                      /rpt_family="MER2_type"
     repeat_region    21592. .21714
                      /rpt_family="MIR"
     repeat_region    21716. .21810
                      /rpt_family="(TA)n"
     repeat_region    22082. .22162
                      /rpt_family="L2"
     repeat_region    23320. .23440
                      /rpt_family="MIR"
BASE COUNT     6642 a    4166 c    4567 g    8106 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 9;  Length 23481;
  Best Local Similarity   100.0%;  Pred. No. 77;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4798 ACTTATATAATATGTATAGAA 4818
              |||||||||||||||||||||
Db       4276 ACTTATATAATATGTATAGAA 4256


RESULT 64
AC014548/c
LOCUS       AC014548               33501 bp    DNA     linear   HTG 16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC014548
VERSION     AC014548.1  GI:6436787
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 33501)
```

```
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10213617 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .33501
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
BASE COUNT    10201 a   6334 c   6522 g  10444 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 2;  Length 33501;
  Best Local Similarity   100.0%;  Pred. No. 79;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2371 TAAATAAATAAAAACAAGCAA 2391
              |||||||||||||||||||||
Db       1400 TAAATAAATAAAAACAAGCAA 1380


RESULT 65
AF206304
LOCUS       AF206304               34718 bp    DNA     linear   VRL 02-APR-2000
DEFINITION  Human herpesvirus 3 ORF49, ORF50, ORF51, ORF52, ORF53, ORF54,
            ORF55, ORF56, ORF57, ORF58, ORF59, ORF60, ORF61, ORF62, ORF63,
            ORF64, ORF65, ORF66, ORF67, and ORF68 genes, complete cds.
ACCESSION   AF206304
VERSION     AF206304.1  GI:7385021
KEYWORDS    .
SOURCE      Human herpesvirus 3
  ORGANISM  Human herpesvirus 3
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Alphaherpesvirinae; Varicellovirus.
REFERENCE   1  (bases 1 to 34718)
  AUTHORS   Argaw,T., Cohen,J.I., Klutch,M., Lekstrom,K., Yoshikawa,T.,
            Asano,Y. and Krause,P.R.
  TITLE     Nucleotide sequences that distinguish Oka vaccine from parental Oka
            and other varicella-zoster virus isolates
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 34718)
  AUTHORS   Argaw,T., Cohen,J.I., Klutch,M., Lekstrom,K., Yoshikawa,T.,
            Asano,Y. and Krause,P.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-1999) Division of Viral Products, FDA/CBER,
            29A/1C16, HFM-457, 29 Lincoln Drive, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .34718
                     /organism="Human herpesvirus 3"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10335"
     CDS             1258. .1503
                     /codon_start=1
                     /product="ORF49"
                     /protein_id="AAF61650.1"
                     /db_xref="GI:7385022"
                     /translation="MGQSSSSGRGGICGLCKRYNELVTCNGETVALNSEFFEDFDFDE
                     NVTEDADKSTQRRPRVIDVTPKRKPSGKSSHSKCAKC"
     CDS             complement(1597. .2904)
                     /codon_start=1
                     /product="ORF50"
                     /protein_id="AAF61651.1"
                     /db_xref="GI:7385023"
                     /translation="MGTQKKGPRSEKVSPYDTTTPEVEALDHQMDTLNWRIWIIQVMM
                     FTLGAVMLLATLIAASSEYTGIPCFYAAVVDYELFNATLDGGVWSGNRGGYSAPVLFL
                     EPHSVVAFTYYTALTAMAMAVYTLITAAIIHRETKNQRVRQSSGVAWLVVDPTTLFWG
                     LLSLWLLNAVVLLLAYKQIGVAATLYLGHFATSVIFTTYFCGRGKLDETNIKAVANLR
                     QQSVFLYRLAGPTRAVFVNLMAALMAICILFVSLMLELVVANHLHTGLWSSVSVAMST
                     FSTLSVVYLIVSELILAHYIHVLIGPSLGTLVACATLGTAAHSYMDRLYDPISVQSPR
                     LIPTTRGTLACLAVFSVVMLLLRLMRAYVYHRQKRSRFYGAVRRVPERVRGYIRKVKP
                     AHRNSRRTNYPSQGYGYVYENDSTYETDREDELLYERSNSGWE"
     CDS             2903. .5410
                     /codon_start=1
                     /product="ORF51"
                     /protein_id="AAF61652.1"
                     /db_xref="GI:7385024"
                     /translation="MSPNTGESNAAVYASSTQLARALYGGDLVSWIKHTHPGISLELQ
                     LDVPVKLIKPGMSQTRPVTVVRAPMGSGKTTALLEWLQHALKADISVLVVSCRRSFTQ
                     TLIQRFNDAGLSGFVTYLTSETYIMGFKRLIVQLESLHRVSSEAIDSYDVLILDEVMS
                     VIGQLYSPTMRRLSAVDSLLYRLLNRCSQIIAMDATVNSQFIDLISGLRGDENIHTIV
                     CTYAGVGFSGRTCTILRDMGIDTLVRVIKRSPEHEDVRTIHQLRGTFFDELALRLQCG
                     HNICIFSSTLSFSELVAQFCAIFTDSILILNSTRPLCNVNEWKHFRVLVYTTVVTVGL
                     SFDMAHFHSMFAYIKPMSYGPDMVSVYQSLGRVRLLLLNEVLMYVDGSRTRCGPLFSP
                     MLLNFTIANKFQWFPTHTQITNKLCCAFRQRCANAFTRSNTHLFSRFKYKHLFERCSL
                     WSLADSINILQTLLASNQILVVLDGMGPITDVSPVQFCAFIHDLRHSANAVASCMRSL
                     RQDNDSCLTDFGPSGFMADNITAFMEKYLMESINTEEQIKVFKALACPIEQPRLVNTA
                     ILGACIRIPEALEAFDVFQKIYTHYASGWFPVLDKTGEFSIATITTAPNLTTHWELFR
                     RCAYIAKTLKWNPSTEGCVTQVLDTDINTLFNQHGDSLAQLIFEVMRCNVTDAKIILN
                     RPVWRTTGFLDGCHNQCFRPIPTKHEYNIALFRLIWEQLFGARVTKSTQTFPGSTRVK
                     NLKKKDLETLLDSINVDRSACRTYRQLYNLLMSHRHSFSQQRYKITAPAWARHVYFQA
                     HQMHLAPHAEAMLQLALSELSPGSWPRINGAVNFESL"
     CDS             5515. .7830
                     /codon_start=1
                     /product="ORF52"
                     /protein_id="AAF61653.1"
                     /db_xref="GI:7385025"
                     /translation="MDATQITLVRESGHICAASIYTSWTQSGQLTQNGLSVLYYLLCK
                     NSCGKYVPKFAEITVQQEDLCRYSRHGGSVSAATFASICRAASSAALDAWPLEPLGNA
                     DTWRCLHGTALATLRRVLGFKSFYSPVTFETDTNTGLLLKTIPDEHALNNDNTPSTGV
                     LRANFPVAIDVSAVSACNAHTQGTSLAYARLTALKSNGDTQQQTPLDVEVITPKAYIR
                     RKYKSTFSPPIERERGQTSDLFNLEERRLVLSGNRAIVVRVLLPCYFDCLTTDSTVTSS
                     LSILATYRLWYAAAFGKPGVVRPIFAYLGPELNPKGEDRDYFCTVGFPGWTTLRTQTP
                     AVESIRTATEMYMETDGLWPVTGIQAFHYLAPWGQHPPLPPRVQDLIGQIPQDTGHAD
                     ATVNWDAGRISTVFKQPVQLQDRWMAKFDFSAFFPTIYCAMFPMHFRLGKIVLARMRR
                     GMGCLKPALVSFFGGLRHILPSIYKAIIFIANEISLCVEQTALEQGFAICTYIKDGFW
                     GIFTDLHTRNVCSDQARCSALNLAAACERAVTGLLRIQLGLNFTPAMEPVLRVEGVYT
                     HAFTWCTTGSWLWNLQTNTPPDLVGVPWRSQAARDLKERLSGLLCTATKIRERIQENC
                     IWDRVLYDIWAGQVVEAARKTYVDFFEHVFDRRYTPVYWSLQEQNSETKAIPASYLTY
                     GHMQDKDYKPRQIIMVRNPNPHGPPTVVYWELLPSCACIPPIDCAAHLKPLIHTFVTI
                     INHLLDAHNDFSSPSLKFTDDPLASYNFLFL"
     CDS             complement(7877. .8872)
                     /codon_start=1
                     /product="ORF53"
                     /protein_id="AAF61654.1"
                     /db_xref="GI:7385026"
                     /translation="MQRIRPYWIKFEQTGGAGMADGMSGINIPSILGCSVTIDNLLTR
                     AEEGLDVSDVIEDLRIQAIPRFVCEAREVTGLKPRFLANSVVSLRVKPEHQETVLVVL
                     NGDSSEVSCDRYYMECVTQPAFRGFIFSVLTAVEDRVYTVGVPPRLLIYRMTLFRPDN
                     VLDFTLCVILMYLEGIGPSGASPSLFVQLSVYLRRVECQIGPLEKMRRFLYEGVLWLL
                     NTLMYVVDNNPFTKTRVLPHYMFVKLLNPQPGTAPNIIKAIYSCGVGQRFDLPHGTPP
                     CPDGVVQVPPGLLNGPLRDSEYQKSVYFWWLNRTMVTPKNVQLFETYKNSPRVVK"
     CDS             complement(8697. .11006)
                     /codon_start=1
                     /product="ORF54"
                     /protein_id="AAF61655.1"
                     /db_xref="GI:7385027"
                     /translation="MAEITSLFNNSSGSEEKRIASSVSIDQGLNGSNPNDQYKNMFDI
                     YWNEYAPDIGFCTFPEEDGWMLIHPTTQSMLFRKILAGDFGYTDGQGIYSAVRSTETV
                     IRQVQATVLMNALDATRYEDLAADWEHHIQQCNLHAGALAERYGLCGESEAVRLAHQV
                     FETWRQTLQSSLLEFLRGITGCLYTSGLNGRVGFAKYVDWIACVGIVPVVRKVRSEQN
                     GTPAPLNTYMGQAAELSQMLKVADATLARGAAVVTSLVECMQNVAIMDYDRTRLYYNY
                     NRRLIMAKDDVTGMKGECLVVWPPVVCGEGVVFDSPLQRLSGEVLACYALREHARVCQ
                     VLNTAPLRVLIGRRNEDDRSHSTRAVDRIMGENDTTRAGSAASRLVKLIVNLKNMRHV
                     GDITETVRSYLEETGNHILEGSGSVDTSQPGFGKANQSFNGGAMSGTTNVQSAFKTSV
                     VNSINGMLEGYVNNLFKTIEGLKDVNSDLTERLQFKEGELKRLREERVKIKPSKGSHI
                     TMAEETRIADLNHEVIDLTGIIGDDAYIANSFQSRYIPPYGDDIKRLSELWKQELVRC
                     FKLHRVNNNQGQEISVSYSNASISLLVAPYFSFILRATRLGFLVTQSEVHRSEEELCQ
                     AIFKKARTESYLSQIRILYEMQVRAEVIKRGPRRTPSPSWGLPDPTEDDERIPEPNKI
                     NNQYMHVGYKNLSHFMKGHPPERLRVHKVNAADSTLLDKIRANRRRGDGRWDVRNKYT
                     QHFRLQRNDRQLTNTSRRGVGCERRDRRS"
```

```
CDS             11018. .13663
                /codon_start=1
                /product="ORF55"
                /protein_id="AAF61656.1"
                /db_xref="GI:7385028"
                /translation="MKRSISVDSSSPKNVFNPETPNGFDDSVYLNFTSMHSIQPILSR
                IRELAAITIPKERVPRLCWFKQLLELQAPPEMQRNELPFSVYLISGNAGSGKSTCIQT
                LNEAIDCIITGSTRVAAQNVHAKLSTAYASRPINTIFHEFGFRGNHIQAQLGRYAYNW
                TTTPPSIEDLQKRDIVYYWEVLIDITKRVFQMGDDGRGGTSTFKTLWAIERLLNKPTG
                SMSGTAFIACGSLPAFTRSNVIVIDEAGLLGRHILTAVVYCWWLLNAIYQSPQYINGR
                KPVIVCVGSPTQTDSLESHFQHDMQRSHVTPSENILTYIICNQTLRQYTNISHNWAIF
                INNKRCQEDDFGNLLKTLEYGLPITEAHARLVDTFVVPASYINNPANLPGWTRLYSSH
                KEVSAYMSKLHAHLKLSKNDHFSVFALPTYTFIRLTAFDEYRKLTGQPGLSVEHWIRA
                NSGRLHNYSQSRDHDMGTVKYETHSNRDLIVARTDITYVLNTLVVVTTRLRKLVIGFS
                GTFQSFAKVLRDDSFVKARGETSIEYAYRFLSNLIFGGLINFYNFLLNKNLHPDKVSL
                AYKRLAALTLELLSGTNKTPLHEAAVNGAGAGIDCDGAATSADKAFCFTKAPESKVTA
                SIPEDPDDVIFTALNDEVIDLVYCQYEFSYPKSSNEVHAQFLLMKAIYDGRYAILAEL
                FESSFTTAPFSAYVDNVNFNGSELLIGNVRGGLLSLALQTDTYTLLGYTFAPVPVFVE
                ELTRKKLYRETTEMLYALHVPLMVLQDQHGFVSIVNANVCEFTESIEDAELAMATTVD
                YGLSSKLAMTIARSQGLSLEKVAICFTADKLRLNSVYVAMSRTVSSRFLKMNLNPLRE
                RYEKSAEISDHILAALRDPNVHVVY"
CDS             13590. .14321
                /codon_start=1
                /product="ORF56"
                /protein_id="AAF61657.1"
                /db_xref="GI:7385029"
                /translation="MKNPQKLAITFLPLYVIPTYTLCIKALYKNTHAGLLFSFLGFVL
                NTPAMSISGPPTTFILYRLHGVRRVLHWTLPDHEQTLYAFTGGSRSMAVKTDARCDTM
                SGGMIVLQHTHTVTLLTIDCSTDFSSYAFTHRDFHLQDKPHATFAMPFMSWVGSDPTS
                QLYSNVGGVLSVITEDDLSMCISIVIYGLRVNRPDDQTTPTPTPHQYTSQRRQPETNC
                PSPQPAFFTSDDDVLSLILRDAANA"
CDS             complement(14430. .14645)
                /codon_start=1
                /product="ORF57"
                /protein_id="AAF61658.1"
                /db_xref="GI:7385030"
                /translation="MDVRERNVFGNASVATPGEHQKFVRELILSGHNNVVLQTYTGKW
                SDCRKHGKSVMYNTGEARHPTCKAHQR"
CDS             complement(14626. .15291)
                /codon_start=1
                /product="ORF58"
                /protein_id="AAF61659.1"
                /db_xref="GI:7385031"
                /translation="MFSELPPSVPTALLQWGWGLHRGPCSIPNFKQVASQHSVQTDFT
                ENSVDANENFPIGHAGCIEKTKDDYVPFDTLFMVSSIDELGRRQLTDTIRRSLIMNAC
                EITVACTKTAAFSGRGVSRQKHVTLSKNKFNPSSHKSLQMFVLCQKTHAPRVRNLLYE
                SIRARRPRRYYTRSTDGKSRPLVPVFVYEFTALDRVLLHKENTLTDQPINTENSGHGR
                TRT"
CDS             complement(15321. .16238)
                /codon_start=1
                /product="ORF59"
                /protein_id="AAF61660.1"
                /db_xref="GI:7385032"
                /translation="MDVSGEPTVCSNAYANEMKLSDSKDIYVLAHPVTKKTRKRPRGP
                PLGVKLDPPTFKLNNMSHHYDTETFTPVSSQLDSVEVFSKFNISPEWYDLLSDELKEP
                YAKGIFLEYNRLLNSGEEILPSTGDIFAWTRFCGPQSIRVVIIGQDPYPTAGHAHGLA
                FSVKRGITPPSSLKNIFAALMESYPNMTPPTHGCLESWARQGVLLLNTTLTVRRGTPG
                SHVYLGWGRLVQRVLQRLCENRTGLVFMLWGAHAQKTTQPNSRCHLVLTHAHPSPLSR
                VPFRNCRHFVQANEYFTRKGEPEIDWSVI"
CDS             complement(16189. .16671)
                /codon_start=1
                /product="ORF60"
                /protein_id="AAF61661.1"
                /db_xref="GI:7385033"
                /translation="MASHKWLLQMIVFLKTITIAYCLHLQDDTPLFFGAKPLSDVSLI
                ITEPCVSSVYEAWDYAAPPVSNLSEALSGIVVKTKCPVPEVILWFKDKQMAYWTNPYV
                TLKGLTQSVGEEHKSGDIRDALLDALSGVWVDSTPSSTNIPENGCVWGADRLFQRVCQ
                "
CDS             complement(18216. .19619)
                /codon_start=1
                /product="ORF61"
                /protein_id="AAF61662.1"
                /db_xref="GI:7385034"
                /translation="MDTILAGGSGTSDASDNTCTICMSTVSDLGKTMPCLHDFCFVCI
                RAWTSTSVQCPLCRCPVQSILHKIVSDTSYKEYEVHPSDDDGFSEPSFEDSIDILPGD
                VIDLLPPSPGPSRESIQQPTSRSSREPIQSPNPGPLQSSAREPTAESPSDSQQDSIQP
                PTRDSSPGVTKTCSTASFLRKVFFKDQPAVRSATPVVYGSIESAQQPRTGGQDYRDRP
                VSVGINQDPRTMDRLPFRATDRGTEGNARFPCYMQPLLGWLDDQLAELYQPEIVEPTK
                MLILNYIGIYGRDEAGLKTSLRCLLHDSTGPFVTNMLFLLDRCTDPTRLTMQTWTWKD
                TAIQLITGPIVRPETTSTGETSRGDERDTRLVNTPQKVRLFSVLPGIKPGSARGAKRR
                LFHTGRDVKRCLTIDLTSESDSACKGSKTRKVASPQGESNTPSTSGSTSGSLKHLTKK
                SSAGKAGKGIPNKMKKS"
CDS             complement(20342. .24274)

  Query Match             0.2%;  Score 21;  DB 14;  Length 34718;
  Best Local Similarity   100.0%;  Pred. No. 79;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        711 CATAGTATGCGTCGGTTCGCC 731
              |||||||||||||||||||||
Db      11854 CATAGTATGCGTCGGTTCGCC 11874


RESULT 66
AC104800/c
LOCUS       AC104800               54618 bp    DNA     linear   PRI 12-JUN-2002
DEFINITION  Homo sapiens BAC clone RP11-350L15 from 4, complete sequence.
ACCESSION   AC104800
VERSION     AC104800.5  GI:21263351
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 54618)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 54618)
  AUTHORS   Tomlinson,C., Meyer,R. and Creason,K.
  TITLE     The sequence of Homo sapiens BAC clone RP11-350L15
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 54618)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 54618)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 54618)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 54618)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On May 30, 2002 this sequence version replaced gi:20377064.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0350L15
            --------------
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-357L18, 2000 bp overlap; the clone sequenced to the right is RP11-568O24, 2000 bp overlap. Actual start of this clone is at base position 9509 of RP11-357L18; actual end is at base position 77842 of RP11-568O24.

```
FEATURES             Location/Qualifiers
     source          1. .54618
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-350L15"
                     /clone_lib="RPCI-11"
     repeat_region   30. .251
                     /rpt_family="L2"
     repeat_region   369. .497
                     /rpt_family="MIR"
     repeat_region   812. .1108
                     /rpt_family="Alu"
     repeat_region   1332. .1767
                     /rpt_family="ERV1"
     repeat_region   2574. .2655
                     /rpt_family="L1"
     repeat_region   2656. .2685
                     /rpt_family="AT_rich"
     repeat_region   2690. .3097
                     /rpt_family="L1"
     repeat_region   3086. .3173
                     /rpt_family="L1"
     repeat_region   3141. .3564
                     /rpt_family="L1"
     repeat_region   3565. .5771
                     /rpt_family="L1"
     repeat_region   5749. .5940
                     /rpt_family="L1"
     repeat_region   6342. .6465
                     /rpt_family="MaLR"
     repeat_region   7644. .7655
                     /rpt_family="L1"
     repeat_region   7656. .8022
                     /rpt_family="L1"
     repeat_region   8023. .8286
                     /rpt_family="L1"
     repeat_region   8287. .8313
                     /rpt_family="(TTG)n"
     repeat_region   8314. .8953
                     /rpt_family="L1"
     repeat_region   9743. .9823
                     /rpt_family="MaLR"
     repeat_region   10794. .10837
                     /rpt_family="AT_rich"
     repeat_region   11545. .11873
                     /rpt_family="L1"
     repeat_region   11900. .12152
                     /rpt_family="Alu"
     repeat_region   12376. .12423
                     /rpt_family="(TATG)n"
     repeat_region   12484. .12550
                     /rpt_family="L1"
     repeat_region   12551. .12575
                     /rpt_family="(CAAAA)n"
     repeat_region   12576. .13165
                     /rpt_family="L1"
     repeat_region   13184. .13224
                     /rpt_family="L1"
     repeat_region   13290. .13448
                     /rpt_family="Alu"
     repeat_region   13506. .13531
                     /rpt_family="AT_rich"
     repeat_region   15831. .16294
                     /rpt_family="L1"
     repeat_region   16290. .16489
                     /rpt_family="L1"
     repeat_region   16568. .16735
                     /rpt_family="(TA)n"
     repeat_region   16731. .16911
                     /rpt_family="(TTATA)n"
     repeat_region   16925. .17098
                     /rpt_family="(TTATA)n"
     repeat_region   17099. .17200
                     /rpt_family="(TA)n"
     repeat_region   17204. .17654
                     /rpt_family="(TTATA)n"
     repeat_region   17659. .18080
                     /rpt_family="(TATAA)n"
     repeat_region   18086. .18245
                     /rpt_family="(TTATA)n"
     repeat_region   18257. .18488
                     /rpt_family="(TTATA)n"
     repeat_region   18493. .18558
                     /rpt_family="(TA)n"
     repeat_region   18603. .18732
                     /rpt_family="(TA)n"
     repeat_region   18765. .19018
                     /rpt_family="L1"
     repeat_region   19548. .20379
                     /rpt_family="L1"
     repeat_region   20463. .20486
                     /rpt_family="AT_rich"
     repeat_region   20502. .20527
                     /rpt_family="AT_rich"
     repeat_region   20642. .20820
                     /rpt_family="L1"
     repeat_region   20950. .21613
                     /rpt_family="L1"
     repeat_region   21614. .21923
                     /rpt_family="Alu"
     repeat_region   21924. .22811
                     /rpt_family="L1"
     repeat_region   22785. .23287
                     /rpt_family="L1"
     repeat_region   23304. .23796
                     /rpt_family="L1"
     repeat_region   23797. .24104
                     /rpt_family="Alu"
```

```
     repeat_region   24105. .24151
                     /rpt_family="L1"
     repeat_region   24145. .24914

  Query Match              0.2%;  Score 21;  DB 9;  Length 54618;
  Best Local Similarity   100.0%;  Pred. No. 81;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1369 ACAAAAAAATGCATATGTGGG 1389
              |||||||||||||||||||||
Db      28914 ACAAAAAAATGCATATGTGGG 28894


RESULT 67
AC101569
LOCUS       AC101569               68882 bp    DNA     linear   HTG 23-NOV-2001
DEFINITION  Mus musculus clone RP23-196G11, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC101569
VERSION     AC101569.1  GI:17060344
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 68882)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-196G11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 68882)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
              Center project name: L17007
              Center clone name: 196_G_11
            --------------
            * NOTE: This record contains 88 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assembled that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      741: contig of 741 bp in length
            *      742      841: gap of 100 bp
            *      842     1524: contig of 683 bp in length
            *     1525     1624: gap of 100 bp
            *     1625     2306: contig of 682 bp in length
            *     2307     2406: gap of 100 bp
            *     2407     3096: contig of 690 bp in length
            *     3097     3196: gap of 100 bp
            *     3197     3870: contig of 674 bp in length
            *     3871     3970: gap of 100 bp
            *     3971     4656: contig of 686 bp in length
            *     4657     4756: gap of 100 bp
            *     4757     5443: contig of 687 bp in length
            *     5444     5543: gap of 100 bp
            *     5544     6228: contig of 685 bp in length
            *     6229     6328: gap of 100 bp
            *     6329     7026: contig of 698 bp in length
            *     7027     7126: gap of 100 bp
            *     7127     7827: contig of 701 bp in length
            *     7828     7927: gap of 100 bp
            *     7928     8605: contig of 678 bp in length
            *     8606     8705: gap of 100 bp
            *     8706     9384: contig of 679 bp in length
            *     9385     9484: gap of 100 bp
            *     9485    10165: contig of 681 bp in length
            *    10166    10265: gap of 100 bp
            *    10266    10949: contig of 684 bp in length
            *    10950    11049: gap of 100 bp
            *    11050    11731: contig of 682 bp in length
            *    11732    11831: gap of 100 bp
            *    11832    12501: contig of 670 bp in length
            *    12502    12601: gap of 100 bp
            *    12602    13286: contig of 685 bp in length
            *    13287    13386: gap of 100 bp
            *    13387    14084: contig of 698 bp in length
            *    14085    14184: gap of 100 bp
            *    14185    14884: contig of 700 bp in length
            *    14885    14984: gap of 100 bp
            *    14985    15664: contig of 680 bp in length
            *    15665    15764: gap of 100 bp
            *    15765    16402: contig of 638 bp in length
            *    16403    16502: gap of 100 bp
            *    16503    17187: contig of 685 bp in length
            *    17188    17287: gap of 100 bp
            *    17288    17962: contig of 675 bp in length
            *    17963    18062: gap of 100 bp
            *    18063    18759: contig of 697 bp in length
            *    18760    18859: gap of 100 bp
            *    18860    19543: contig of 684 bp in length
            *    19544    19643: gap of 100 bp
            *    19644    20325: contig of 682 bp in length
            *    20326    20425: gap of 100 bp
            *    20426    21106: contig of 681 bp in length
            *    21107    21206: gap of 100 bp
            *    21207    21905: contig of 699 bp in length
            *    21906    22005: gap of 100 bp
            *    22006    22692: contig of 687 bp in length
            *    22693    22792: gap of 100 bp
            *    22793    23495: contig of 703 bp in length
            *    23496    23595: gap of 100 bp
            *    23596    24284: contig of 689 bp in length
            *    24285    24384: gap of 100 bp
            *    24385    25079: contig of 695 bp in length
            *    25080    25179: gap of 100 bp
            *    25180    25857: contig of 678 bp in length
            *    25858    25957: gap of 100 bp
            *    25958    26611: contig of 654 bp in length
            *    26612    26711: gap of 100 bp
            *    26712    27410: contig of 699 bp in length
```

```
        *    27411    27510: gap of 100 bp
        *    27511    28207: contig of 697 bp in length
        *    28208    28307: gap of 100 bp
        *    28308    28995: contig of 688 bp in length
        *    28996    29095: gap of 100 bp
        *    29096    29771: contig of 676 bp in length
        *    29772    29871: gap of 100 bp
        *    29872    30567: contig of 696 bp in length
        *    30568    30667: gap of 100 bp
        *    30668    31359: contig of 692 bp in length
        *    31360    31459: gap of 100 bp
        *    31460    32143: contig of 684 bp in length
        *    32144    32243: gap of 100 bp
        *    32244    32920: contig of 677 bp in length
        *    32921    33020: gap of 100 bp
        *    33021    33710: contig of 690 bp in length
        *    33711    33810: gap of 100 bp
        *    33811    34485: contig of 675 bp in length
        *    34486    34585: gap of 100 bp
        *    34586    35274: contig of 689 bp in length
        *    35275    35374: gap of 100 bp
        *    35375    36053: contig of 679 bp in length
        *    36054    36153: gap of 100 bp
        *    36154    36813: contig of 660 bp in length
        *    36814    36913: gap of 100 bp
        *    36914    37594: contig of 681 bp in length
        *    37595    37694: gap of 100 bp
        *    37695    38385: contig of 691 bp in length
        *    38386    38485: gap of 100 bp
        *    38486    39178: contig of 693 bp in length
        *    39179    39278: gap of 100 bp
        *    39279    39965: contig of 687 bp in length
        *    39966    40065: gap of 100 bp
        *    40066    40751: contig of 686 bp in length
        *    40752    40851: gap of 100 bp
        *    40852    41547: contig of 696 bp in length
        *    41548    41647: gap of 100 bp
        *    41648    42316: contig of 669 bp in length
        *    42317    42416: gap of 100 bp
        *    42417    43072: contig of 656 bp in length
        *    43073    43172: gap of 100 bp
        *    43173    43860: contig of 688 bp in length
        *    43861    43960: gap of 100 bp
        *    43961    44640: contig of 680 bp in length
        *    44641    44740: gap of 100 bp
        *    44741    45417: contig of 677 bp in length
        *    45418    45517: gap of 100 bp
        *    45518    46276: contig of 759 bp in length
        *    46277    46376: gap of 100 bp
        *    46377    47024: contig of 648 bp in length
        *    47025    47124: gap of 100 bp
        *    47125    47818: contig of 694 bp in length
        *    47819    47918: gap of 100 bp
        *    47919    48612: contig of 694 bp in length
        *    48613    48712: gap of 100 bp
        *    48713    49390: contig of 678 bp in length
        *    49391    49490: gap of 100 bp
        *    49491    50168: contig of 678 bp in length
        *    50169    50268: gap of 100 bp
        *    50269    50947: contig of 679 bp in length
        *    50948    51047: gap of 100 bp
        *    51048    51735: contig of 688 bp in length
        *    51736    51835: gap of 100 bp
        *    51836    52510: contig of 675 bp in length
        *    52511    52610: gap of 100 bp
        *    52611    53291: contig of 681 bp in length
        *    53292    53391: gap of 100 bp
        *    53392    54029: contig of 638 bp in length
        *    54030    54129: gap of 100 bp

Query Match             0.2%;  Score 21;  DB 2;  Length 68882;
Best Local Similarity   100.0%;  Pred. No. 82;
Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2545 GTCTCTTAGGGCCCTGGGTTG 2565
              |||||||||||||||||||||
Db       1052 GTCTCTTAGGGCCCTGGGTTG 1072


RESULT 68
AC101569/c
LOCUS       AC101569               68882 bp    DNA     linear   HTG 23-NOV-2001
DEFINITION  Mus musculus clone RP23-196G11, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC101569
VERSION     AC101569.1  GI:17060344
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 68882)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-196G11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 68882)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L17007
                Center clone name: 196_G_11
            --------------.
            * NOTE: This record contains 88 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      741: contig of 741 bp in length
            *      742      841: gap of 100 bp
            *      842     1524: contig of 683 bp in length
```

```
*     1525    1624: gap of 100 bp
*     1625    2306: contig of 682 bp in length
*     2307    2406: gap of 100 bp
*     2407    3096: contig of 690 bp in length
*     3097    3196: gap of 100 bp
*     3197    3870: contig of 674 bp in length
*     3871    3970: gap of 100 bp
*     3971    4656: contig of 686 bp in length
*     4657    4756: gap of 100 bp
*     4757    5443: contig of 687 bp in length
*     5444    5543: gap of 100 bp
*     5544    6228: contig of 685 bp in length
*     6229    6328: gap of 100 bp
*     6329    7026: contig of 698 bp in length
*     7027    7126: gap of 100 bp
*     7127    7827: contig of 701 bp in length
*     7828    7927: gap of 100 bp
*     7928    8605: contig of 678 bp in length
*     8606    8705: gap of 100 bp
*     8706    9384: contig of 679 bp in length
*     9385    9484: gap of 100 bp
*     9485   10165: contig of 681 bp in length
*    10166   10265: gap of 100 bp
*    10266   10949: contig of 684 bp in length
*    10950   11049: gap of 100 bp
*    11050   11731: contig of 682 bp in length
*    11732   11831: gap of 100 bp
*    11832   12501: contig of 670 bp in length
*    12502   12601: gap of 100 bp
*    12602   13286: contig of 685 bp in length
*    13287   13386: gap of 100 bp
*    13387   14084: contig of 698 bp in length
*    14085   14184: gap of 100 bp
*    14185   14884: contig of 700 bp in length
*    14885   14984: gap of 100 bp
*    14985   15664: contig of 680 bp in length
*    15665   15764: gap of 100 bp
*    15765   16402: contig of 638 bp in length
*    16403   16502: gap of 100 bp
*    16503   17187: contig of 685 bp in length
*    17188   17287: gap of 100 bp
*    17288   17962: contig of 675 bp in length
*    17963   18062: gap of 100 bp
*    18063   18759: contig of 697 bp in length
*    18760   18859: gap of 100 bp
*    18860   19543: contig of 684 bp in length
*    19544   19643: gap of 100 bp
*    19644   20325: contig of 682 bp in length
*    20326   20425: gap of 100 bp
*    20426   21106: contig of 681 bp in length
*    21107   21206: gap of 100 bp
*    21207   21905: contig of 699 bp in length
*    21906   22005: gap of 100 bp
*    22006   22692: contig of 687 bp in length
*    22693   22792: gap of 100 bp
*    22793   23495: contig of 703 bp in length
*    23496   23595: gap of 100 bp
*    23596   24284: contig of 689 bp in length
*    24285   24384: gap of 100 bp
*    24385   25079: contig of 695 bp in length
*    25080   25179: gap of 100 bp
*    25180   25857: contig of 678 bp in length
*    25858   25957: gap of 100 bp
*    25958   26611: contig of 654 bp in length
*    26612   26711: gap of 100 bp
*    26712   27410: contig of 699 bp in length
*    27411   27510: gap of 100 bp
*    27511   28207: contig of 697 bp in length
*    28208   28307: gap of 100 bp
*    28308   28995: contig of 688 bp in length
*    28996   29095: gap of 100 bp
*    29096   29771: contig of 676 bp in length
*    29772   29871: gap of 100 bp
*    29872   30567: contig of 696 bp in length
*    30568   30667: gap of 100 bp
*    30668   31359: contig of 692 bp in length
*    31360   31459: gap of 100 bp
*    31460   32143: contig of 684 bp in length
*    32144   32243: gap of 100 bp
*    32244   32920: contig of 677 bp in length
*    32921   33020: gap of 100 bp
*    33021   33710: contig of 690 bp in length
*    33711   33810: gap of 100 bp
*    33811   34485: contig of 675 bp in length
*    34486   34585: gap of 100 bp
*    34586   35274: contig of 689 bp in length
*    35275   35374: gap of 100 bp
*    35375   36053: contig of 679 bp in length
*    36054   36153: gap of 100 bp
*    36154   36813: contig of 660 bp in length
*    36814   36913: gap of 100 bp
*    36914   37594: contig of 681 bp in length
*    37595   37694: gap of 100 bp
*    37695   38385: contig of 691 bp in length
*    38386   38485: gap of 100 bp
*    38486   39178: contig of 693 bp in length
*    39179   39278: gap of 100 bp
*    39279   39965: contig of 687 bp in length
*    39966   40065: gap of 100 bp
*    40066   40751: contig of 686 bp in length
*    40752   40851: gap of 100 bp
*    40852   41547: contig of 696 bp in length
*    41548   41647: gap of 100 bp
*    41648   42316: contig of 669 bp in length
*    42317   42416: gap of 100 bp
*    42417   43072: contig of 656 bp in length
*    43073   43172: gap of 100 bp
*    43173   43860: contig of 688 bp in length
*    43861   43960: gap of 100 bp
*    43961   44640: contig of 680 bp in length
*    44641   44740: gap of 100 bp
*    44741   45417: contig of 677 bp in length
*    45418   45517: gap of 100 bp
*    45518   46276: contig of 759 bp in length
*    46277   46376: gap of 100 bp
*    46377   47024: contig of 648 bp in length
*    47025   47124: gap of 100 bp
*    47125   47818: contig of 694 bp in length
*    47819   47918: gap of 100 bp
*    47919   48612: contig of 694 bp in length
*    48613   48712: gap of 100 bp
*    48713   49390: contig of 678 bp in length
*    49391   49490: gap of 100 bp
*    49491   50168: contig of 678 bp in length
*    50169   50268: gap of 100 bp
*    50269   50947: contig of 679 bp in length
*    50948   51047: gap of 100 bp
*    51048   51735: contig of 688 bp in length
*    51736   51835: gap of 100 bp
*    51836   52510: contig of 675 bp in length
*    52511   52610: gap of 100 bp
*    52611   53291: contig of 681 bp in length
*    53292   53391: gap of 100 bp
*    53392   54029: contig of 638 bp in length
*    54030   54129: gap of 100 bp

  Query Match              0.2%;  Score 21;  DB 2;  Length 68882;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2545 GTCTCTTAGGGCCCTGGGTTG 2565
              |||||||||||||||||||||
Db       4094 GTCTCTTAGGGCCCTGGGTTG 4074


RESULT 69
```

```
AL359960/c
LOCUS       AL359960              69113 bp    DNA     linear   PRI 15-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-298K24 on chromosome 10,
            complete sequence.
ACCESSION   AL359960
VERSION     AL359960.23  GI:16972932
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 69113)
  AUTHORS   Howden,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 16, 2001 this sequence version replaced gi:15982053.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-298K24 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-298K24 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP11-563J2 is at 67114 in this sequence.
            The true right end of clone RP11-414H17 is at 2000 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .69113
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-298K24"
                     /clone_lib="RPCI-11.2"
     misc_feature    31391. .31668
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
BASE COUNT    17147 a  15616 c  15801 g  20549 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 9;  Length 69113;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8682 ACACTCCAGAGTCCCGCATGG 8702
             |||||||||||||||||||||
Db      29123 ACACTCCAGAGTCCCGCATGG 29103
```

```
RESULT 70
AL591511
LOCUS       AL591511              74508 bp    DNA     linear   VRT 13-DEC-2002
DEFINITION  Zebrafish DNA sequence from clone BUSM1-1C23 Contains 28 TCR-alpha
            V segments (II-48 to II-63, II-65 to II-76) one of which is a
            pseudogene (II-53) and a CpG island, complete sequence.
ACCESSION   AL591511
VERSION     AL591511.2  GI:15591539
KEYWORDS    HTG; CpG island; T-cell receptor; TCR-alpha; V segment; variable
            region.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 74508)
  AUTHORS   Corby,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zface@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 12, 2001 this sequence version replaced gi:14141560.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This clone was
            isolated from the BUSM1 library (C. Amemiya) and provided by C.
            Amemiya (Amemiya lab, Virginia Mason Research Center, Seattle,
            USA). Tatsuya Ota (Kanagawa, Japan) kindly provided the annotation
            for this clone.
            BUSM1-1C23 is from a Zebrafish PAC library
            VECTOR: pCYPAC-6
            This sequence is the entire insert of clone BUSM1-1C23 The true
            right end of clone BUSM1-18F12 is at 16446 in this sequence.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zface@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            Repeat names beginning 'Dr' were identified by the Recon repeat
            discovery system (Zhirong Bao and Sean Eddy, submitted), and those
            beginning 'drr' were identified by Rick Waterman (Stephen Johnson
            lab, WashU).  For further information see
            http://www/Projects/D_rerio/fishmask.shtml.
FEATURES             Location/Qualifiers
     source          1. .74508
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="BUSM1-1C23"
                     /clone_lib="BUSM1"
     gene            1. .55
                     /gene="dZ18F12.36"
     CDS             <1. .>55
                     /gene="dZ18F12.36"
                     /note="according to annotation by Tatsuya Ota
```

```
                identical to SI:dZ18F12.36 due to clone overlap"
                /codon_start=2
                /evidence=not_experimental
                /product="SI:dZ1C23.1 (TCR-alpha V segment II-48)"
                /protein_id="CAD24441.2"
                /db_xref="GI:20196558"
                /db_xref="REMTREMBL:CAD24441"
                /translation="IKDLRVSDSAVYYCALRP"
     repeat_region   69. .82
                /note="2.0 copies 7 mer ACACTCA 28% conserved"
     repeat_region   147. .160
                /note="2.0 copies 7 mer TAACTAG 28% conserved"
     repeat_region   221. .334
                /note="HATN10_DR repeat: matches 337. .455 of consensus"
     repeat_region   339. .362
                /note="3.4 copies 7 mer CTAGTAA 25% conserved"
     repeat_region   340. .364
                /note="2.1 copies 12 mer TAGTAACTAGTT 50% conserved"
     repeat_region   419. .429
                /note="3.7 copies 3 mer TAT 22% conserved"
     repeat_region   448. .472
                /note="3.1 copies 8 mer TATAAATG 32% conserved"
     repeat_region   490. .502
                /note="2.2 copies 6 mer TATTTA 26% conserved"
     repeat_region   582. .592
                /note="2.2 copies 5 mer TAAAA 22% conserved"
     misc_feature    675. .684
                /note="Inconsistency in the number of bases in
                mononucleotide run between subclones."
     repeat_region   675. .684
                /note="10.0 copies 1 mer A 20% conserved"
     repeat_region   complement(835. .1382)
                /note="HATN12_DR repeat: matches 18. .579 of consensus"
     repeat_region   1559. .1568
                /note="2.5 copies 4 mer AAGA 20% conserved"
     repeat_region   1666. .1768
                /note="HATN10_DR repeat: matches 7. .113 of consensus"
     repeat_region   2014. .2031
                /note="2.6 copies 7 mer AAAATTA 36% conserved"
     repeat_region   complement(2089. .2353)
                /note="DNA21TA1_DR repeat: matches 249. .530 of consensus"
     repeat_region   2351. .2668
                /note="Dr000073 repeat: matches 1. .322 of consensus"
     repeat_region   complement(2666. .2907)
                /note="DNA21TA1_DR repeat: matches 1. .250 of consensus"
     repeat_region   3028. .3248
                /note="HATN10_DR repeat: matches 262. .455 of consensus"
     gene            join(3382. .3430,3508. .3812)
                /gene="SI:dZ1C23.2"
     CDS             join(3382. .3430,3508. .>3812)
                /gene="SI:dZ1C23.2"
                /note="according to annotation by Tatsuya Ota"
                /codon_start=1
                /evidence=not_experimental
                /product="SI:dZ1C23.2 (TCR-alpha V segment II-49)"
                /protein_id="CAD24442.2"
                /db_xref="GI:20196559"
                /db_xref="REMTREMBL:CAD24442"
                /translation="MSNWIRFSLILFSFSLECRAQDSVTQTATVQTATEGQAVTIECT
                YKTNAFPTLYWYQYKANQLPKYMLKRYAGSSDEDDNFKDRFNANLNSSSTSVPLTITD
                VRVSDSAVYYCALQPT"
     repeat_region   3972. .4076
                /note="ANGEL repeat: matches 215. .317 of consensus"
     repeat_region   4191. .4211
                /note="21.0 copies 1 mer A 24% conserved"
     repeat_region   4274. .4778
                /note="DNA21TA1_DR repeat: matches 2. .507 of consensus"
     repeat_region   4853. .4863
                /note="2.2 copies 5 mer GCAGA 22% conserved"
     repeat_region   complement(5016. .5179)
                /note="Dr000032 repeat: matches 213. .403 of consensus"
     repeat_region   5290. .5314
                /note="8.3 copies 3 mer TTA 50% conserved"
     repeat_region   complement(5356. .5446)
                /note="Dr000184 repeat: matches 362. .462 of consensus"
     repeat_region   5366. .5451
                /note="Dr000228 repeat: matches 43. .465 of consensus"
     repeat_region   5452. .5719
                /note="HATN10_DR repeat: matches 171. .445 of consensus"
     repeat_region   5722. .5736
                /note="2.1 copies 7 mer TGTATAC 30% conserved"
     repeat_region   5804. .5818
                /note="15.0 copies 1 mer T 30% conserved"
     gene            join(6019. .6073,6163. .6467)
                /gene="SI:dZ1C23.3"
     CDS             join(6019. .6073,6163. .>6467)
                /gene="SI:dZ1C23.3"
                /note="according to annotation by Tatsuya Ota"
                /codon_start=1
                /evidence=not_experimental
                /product="SI:dZ1C23.3 (TCR-alpha V segment II-50)"
                /protein_id="CAD24443.2"
                /db_xref="GI:20196560"
                /db_xref="REMTREMBL:CAD24443"
                /translation="MLTFAVVLLAEFLSIAFHGVSSAEEIISPNNKTVIAKADEKVLL
                SCSYTGTVYDLYWYRQSPGSAPQFLIMESSGVIIHATPQVPGIKMINKKNTKHVELEI
                FSASVEHSAVYYCAVRPT"
     repeat_region   6711. .6720
                /note="2.5 copies 4 mer ATAA 20% conserved"
     repeat_region   6789. .6799
                /note="2.2 copies 5 mer ATTTT 22% conserved"
     repeat_region   7061. .7434
                /note="TDR18 repeat: matches 153. .568 of consensus"
     gene            join(7635. .7692,7782. .8088)
                /gene="SI:dZ1C23.4"
     CDS             join(7635. .7692,7782. .>8088)
                /gene="SI:dZ1C23.4"
                /note="according to annotation by Tatsuya Ota"
                /codon_start=1
                /evidence=not_experimental
                /product="SI:dZ1C23.4 (TCR-alpha V segment II-51)"
                /protein_id="CAD24444.2"
                /db_xref="GI:20196561"
                /db_xref="REMTREMBL:CAD24444"
                /translation="MITARRCLETLLILFATSMESRAQDSVYQPKTLQSANEEDAAVT
                LDCKYSTSSSNPTLLWYRQDDYGSPKYILLRYKFGDGDEATEFKGRFGSRLDSASKSV
                PLTIQHPHASDSAVYFCALL"
     repeat_region   7890. .7901
                /note="2.4 copies 5 mer CTACT 24% conserved"
     repeat_region   7957. .7970
                /note="2.3 copies 6 mer TGGAGA 28% conserved"
     repeat_region   8373. .8402
                /note="DNA11TA1_DR repeat: matches 174. .203 of consensus"
     repeat_region   8425. .8440
                /note="2.0 copies 8 mer GATTGAAT 32% conserved"
     repeat_region   8504. .8744
                /note="TDR12 repeat: matches 2. .242 of consensus"
     repeat_region   8937. .9045
                /note="LOOPERN1_DR repeat: matches 1. .111 of consensus"
     repeat_region   9071. .9080

  Query Match             0.2%;  Score 21;  DB 5;  Length 74508;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4712 TAATTAAAAAAAATCCACATG 4732
              |||||||||||||||||||||
Db      56078 TAATTAAAAAAAATCCACATG 56098


RESULT 71
AB025615
LOCUS       AB025615               83544 bp    DNA     linear   PLN 27-DEC-2000
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 3, TAC clone:K5K13.
ACCESSION   AB025615 BA000014
VERSION     AB025615.1  GI:4589421
```

```
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (sites)
  AUTHORS   Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E. and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 3. I.
            Sequence features of the regions of 4,504,864 bp covered by sixty
            P1 and TAC clones
  JOURNAL   DNA Res. 7 (2), 131-135 (2000)
  MEDLINE   20277480
   PUBMED   10819329
REFERENCE   2  (bases 1 to 83544)
  AUTHORS   Sato,S., Nakamura,Y., Kaneko,T., Kato,T., Asamizu,E. and Tabata,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=K5K13
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers.  Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/Grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SplicePredictor (Volker Brendel, Stanford University,
            http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St. Louis,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
            This sequence may not be the entire insert of this clone. It may be
            shorter because we remove overlaps between neighboring submissions.
            The 5' clone is MYI13 and the 3' clone is MRI12.
FEATURES             Location/Qualifiers
     source          1. .83544
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="3"
                     /clone="K5K13"
                     /clone_lib="Mitsui TAC"
     CDS             2522. .2917
                     /note="gene_id:K5K13.1
                     unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAA95742.1"
                     /db_xref="GI:7939539"
                     /translation="MYLPSRWPAMSIYHALLESSSLTSYPLPLLTPISLTGESKLPWS
                     STCALNPESTPQSPPFRSPEMLEPPRNPSASRPCHLDFGSDSVWIISDSDNSDERVII
                     SSDYMTIRFGFGLDCVGFRSVSDQFLRIF"
     CDS             join(7732. .7828,7956. .8021,8082. .8195,8616. .8983)
                     /note="gene_id:K5K13.2
                     unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAA95743.1"
                     /db_xref="GI:7939540"
                     /translation="MLRNRELSDKGRKSCSNHLYESMKQRSSESGKEQKKHRKKKTMS
                     MSEKRHHDEKDPRRSLEVAGSSDGWNKTNITCCGFGSGVITGFGLEPPEILWRELQHN
                     FCRNHLKNKKHVFHKSRQPDKRLLSSNQLIRQRALHMVGTEKFEELVVTLDGALFEKY
                     GTMSGDSIELSNIVDMLSNIREKIPNAVRQYRAAENKVSHLEMGLAKAKERYWL"
     CDS             complement(join(9292. .9429,9520. .9603,9688. .9780,
                     9850. .9922,10006. .10130))
                     /note="gene_id:K5K13.3"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="AIG2 protein-like"
                     /protein_id="BAA95744.1"
                     /db_xref="GI:7939541"
                     /translation="MTSSDQSPSHDVFVYGSFQEPAVVNLILECAPVMVSAQLHGYHL
                     YRLKGRLHPCISPSDNGLINGKILTGLTDSQLESLDMIEGTEYVRKTVEVVLTDTLEK
                     KQVETIVWANKDDPNMYGEWDFEEWKRLHMEKFIEAATKFMEWKKNPNGRSREEFEKF
                     VQDDSSPASA"
     CDS             complement(join(17726. .17860,17936. .18019,18107. .18199,
                     18410. .18482,18589. .18713))
                     /note="gene_id:K5K13.4"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="AIG2 protein-like"
                     /protein_id="BAA95745.1"
                     /db_xref="GI:7939542"
                     /translation="MTSSDKSLSHDVFVYGSFQEPSVVNLILECSPVMVPAQLHGYHV
                     YRLKGRLHACISPSENGLINGKILTGLTDSQLENLDMIEGSEYVRKTVEVVLTDTSEK
                     KQVETYVWANKDDPNLYGEWDFEEWRRLHMDKFIEASTKFIEWKKNPDGRSREEFEKF
                     VFDDPPTAA"
     CDS             complement(19678. .25253)
                     /note="gene_id:K5K13.5"
                     /pseudo
                     /codon_start=1
                     /evidence=not_experimental
                     /product="non-LTR retroelement protein"
     CDS             complement(join(25761. .25886,26136. .26219,26319. .26411,
                     26505. .26577,26651. .26772))
                     /note="gene_id:K5K13.6"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="AIG2 protein-like"
                     /protein_id="BAA95746.1"
                     /db_xref="GI:7939543"
                     /translation="MTSSDPQSHNVFVYGSILEPAVAAVILDRTADTVPAVLHGYHRY
                     KLKGLPYPCIVSSDSGKVNGKVITGVSDAELNNFDVIEGNDYERVTVEVVRMDNSEKV
                     KVETYVWVNKDDPRMYGEWDFEEWRVVHAEKFVETFRKMLEWNKNPNGKSMEEAVGSL
                     LSSGD"
     CDS             complement(join(27842. .27936,28031. .28295,28397. .28549))
                     /note="gene_id:K5K13.7"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="replication protein-like"
                     /protein_id="BAA95747.1"
                     /db_xref="GI:7939544"
                     /translation="MMLANPLPFTPLKSLKPYKNAWRIQVKLLHVWRQYSVKVGESIE
                     MILVNEVGDKMYAPVRREQIEKIERCLTEGVWKIITTITLNHTSGQYRISDLKYKIGF
                     VFKTTVSPCDTVFDALFLSLAKFDVILSGSANSNILHDVMGQVVDRSEIKDLNANNKS
                     TKKIDFHLRD"
     CDS             complement(join(30191. .30334,30437. .30589,30666. .30809,
                     31455. .31520))
                     /note="gene_id:K5K13.8
                     unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAA95748.1"
                     /db_xref="GI:7939545"
                     /translation="MDFGLQRDSEMNRLSFCIGFSQSKFGVVICRMKANAGALTNFEL
                     LDFLNSRGASKDTTRVIAPIARSEYKVYDYLVETAASTQTRESVNKSADKCKDFKLAK
                     AEILNIINLWPSSIVELLPIIENLDDREIDTDGILELVKDLLPPLPTAESPKDNDEEE
                     ETENGEQS"
     CDS             complement(join(33437. .33549,33688. .34165,34396. .35003,
                     35079. .35169))
                     /note="gb|AAC79623.2
                     gene_id:K5K13.9
                     similar to unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAA95749.1"
                     /db_xref="GI:7939546"
                     /translation="MEEDNQEIQRDNVGSSSFFKTCFNALNALSGIGILSVPYSLARG
```

```
                     GWLSLSLLLLLAVTAFYTSLLITKCMNADRNIKTYPDIGERAFGRPGRIIVSVFMHLE
                     LYLVTTGFLILEGDNLHNLFPGFTIEMIGLRLNGKQAFMATVAFVIMPTLWWDNLSVL
                     SYVSMSGVLATTVTLGSISWIGAFDGIGFHQKGKLINWSGIPTALSLYAFCYGAHPVL
                     PTLYSSMKSKHQFNNVLLICFILCTIGYTSMAVLGYLMYGSQTLSQITLNLPIHKTSS
                     KVAIYTTLVNPVAKYALMITPTVNTIKDWFPSRYSKKAYLHLLISTFFIISSVVIAET
                     LPFFGYMMSLVGALLSVTVSILLPCLCYLKIFGNYKKIGCETIMLFGMVVMSVFVGVI
                     GTYIAISSSIEVLETFEWLVFYFHQNLSYFSLALHFT"
     CDS             complement(join(37372. .37637,37725. .37804,37892. .37937,
                     38007. .38103,38184. .38266,38346. .38421,38509. .38560,
                     38761. .38887,39124. .39184))
                     /note="gene_id:K5K13.10
                     unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAA95750.1"
                     /db_xref="GI:7939547"
                     /translation="MDSSPVSARFDIFEIYRRFCAIRSGQQLCNKKPCDGEESQRFNL
                     SKEAITQLLYLVENKFQARNSIFDELFKLMSRLDLMVDFTEFTCFYDFVFFMCRENGQ
                     KNITISRAITAWKLVLAGRFRLLNRWCDFIEKNQRHNISEDTWQQVLAFSRCVHENLE
                     GYDSEGAWPVLIDDFVEHMYSILGPNKDTSLFCKCGDTESESCLYQEDEHHKDYRRPH
                     TGLRNIPGLKRKTSKKNDEEEEDEDEEVLETQNSSSLLNFKRIKTSNSPRCSSKSPCS
                     IERSLSQGFASLLSTGDKP"
     CDS             complement(join(42805. .43077,43181. .43894,43977. .44327))
                     /note="gene_id:K5K13.11
                     unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAA95751.1"
                     /db_xref="GI:7939548"
                     /translation="MARVQLLLCFTILFASVTLLDLVSAHLKFKPSLPQIEDPKTVKD
                     VEPYTVKVVMVFVADLEKECPKTNKFKAFFEKLRGFAKYVCPIRRIGQVDYDRDMKAK
                     AGGIFKAISSFAIGKIREEIQEEKQEAIDTFRFMKSVAAKIMGGRKKEENEETMTLTA
                     EQQKEIKEGILKWETIITRITNTMVSSTSTSASSSEESSVGKESETSAKGESETSAKN
                     ESKTSAKGESETSAKGESKTSAKGESETSAKGESETSAKGESETSSSKTAEASQSSSL
                     TVTQVEEETSKDVSTFIMNLEKKCPQKEEFKVFFEQLKGTMIAPRKERKGLFSRIKTA
                     AGKISGAMQVMRSRIGSKSAEVKKNMEAYQEQVMKTLEELDTIHAQIVSQNKGKASLT
                     CTPTQQVQIKQTITKWEQVTTQFVEVAAQSESSASSSSSSSSSSSSSGKMQAK"
     CDS             complement(join(51753. .51877,52193. .52298,52413. .52448))
                     /note="emb|CAB75758.1
                     gene_id:K5K13.12
                     similar to unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAA95752.1"
                     /db_xref="GI:7939549"
                     /translation="MRHSSPAVYIHLVQHMIETCLTFNMSKEECMEALSENANINPII
                     TSTVWKELVKENKDFFETYEQKLMKKESMSEEETNQLIQNIISL"
     CDS             55193. .55777
                     /note="contains similarity to calmodulin
                     gene_id:K5K13.13"

  Query Match             0.2%;  Score 21;  DB 8;  Length 83544;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3924 TCTCTTTAAAAATATTTTTGA 3944
              |||||||||||||||||||||
Db       6218 TCTCTTTAAAAATATTTTTGA 6238


RESULT 72
AL831757/c
LOCUS       AL831757               85948 bp    DNA     linear   ROD 10-SEP-2002
DEFINITION  Mouse DNA sequence from clone RP23-279A21 on chromosome X, complete
            sequence.
ACCESSION   AL831757
VERSION     AL831757.5  GI:22859066
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 85948)
  AUTHORS   Chapman,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 14, 2002 this sequence version replaced gi:22798454.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-279A21 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES              Location/Qualifiers
     source           1. .85948
                      /organism="Mus musculus"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:10090"
                      /chromosome="X"
                      /clone="RP23-279A21"
                      /clone_lib="RPCI-23"
BASE COUNT    27656 a  16297 c  16280 g  25715 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 10;  Length 85948;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2372 AAATAAATAAAAACAAGCAAA 2392
              |||||||||||||||||||||
Db      83003 AAATAAATAAAAACAAGCAAA 82983


RESULT 73
AC002538/c
LOCUS       AC002538               106793 bp    DNA     linear   PRI 31-OCT-1997
DEFINITION  Homo sapiens chromosome 17, clone 138g16, complete sequence.
ACCESSION   AC002538
VERSION     AC002538.1  GI:2580473
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 106793)
  AUTHORS   Birren,B., Fasman,K., McKernan,K., Munro,C., Nusbaum,C.,
            Richardson,P. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone 138g16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 106793)
  AUTHORS   Hawkins,T.L., Birren,B.W., Fasman,K.H., Nussbaum,C., Lander,E.S.,
            McKernan,K., Munro,C., Richardson,P., Barna,N., Chang,A., Cooke,P.,
```

```
            Daly,M.J., Devon,K., Dewar,K., Forrest,C., Gage,D., Geraigery,K.,
            Hagos,B., Harris,K., Huang,J., Hui,L., Jacotot,L., Kirby,A.,
            Lane,M., MacKenzie,J., Marquis,N., McDermott,J., Molla,M.,
            Morrow,J., Nachman,A., Naylor,J., Nusbaum,C., O'Connor,T.,
            Olotu,A., Peterson,K., Reeve,M.P., Roberts,D., Rollins,G.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L.,
            Wilmer,F., Zemtseva,I. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 106793)
  AUTHORS   Birren,B., Fasman,K., McKernan,K., Munro,C., Nusbaum,C.,
            Richardson,P., Lander,E., Baldwin,J., Barna,N., Cantu,C., Chang,A.,
            Cooke,P., Daly,M.J., Devon,K., Dewar,K., DuRette,B., Forrest,C.,
            Gage,D., Gensheimer,S., Geraigery,K., Gilmartin,T., Hagos,B.,
            Halphen,I., Harris,K., Howland,J.C., Huang,J., Hui,L., Jacotot,L.,
            Kirby,A., Lane,M., MacKenzie,J., Marquis,N., McDermott,J.,
            Molla,M., Morrow,J., Nachman,A., Naylor,J., O'Connor,T., Olotu,A.,
            Peterson,K., Roberts,D., Rollins,G., Sarnaik,A., Shiu,P., Shyam,R.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L.,
            Zemtseva,I. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 31, 1997 this sequence version replaced gi:2570238.
            The Staden databases, finishing information, and all
            chromatographic files used in the assembly of this clone are
            available from our anonymous ftp site.

            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             Location/Qualifiers
     source          1. .106793
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /clone="138g16"
                     /clone_lib="Peter de Jong/ human PAC library"
     repeat_region   967. .1268
                     /rpt_family="AluSx"
     repeat_region   complement(1269. .1296)
                     /rpt_family="AT_rich"
     repeat_region   1361. .1659
                     /rpt_family="AluY"
     repeat_region   complement(2382. .2568)
                     /rpt_family="MIR"
     repeat_region   2709. .3071
                     /rpt_family="THE1B"
     repeat_region   4462. .4514
                     /rpt_family="(CA)n"
     repeat_region   complement(5249. .5345)
                     /rpt_family="(GAAA)n"
     repeat_region   complement(6332. .6367)
                     /rpt_family="AT_rich"
     repeat_region   7181. .7333
                     /rpt_family="MIR"
     repeat_region   complement(8026. .8145)
                     /rpt_family="MER5A"
     repeat_region   complement(8146. .8446)
                     /rpt_family="AluY"
     repeat_region   complement(8448. .8525)
                     /rpt_family="MER5A"
     repeat_region   8603. .8636
                     /rpt_family="tRNA-Lys-AAG"
     repeat_region   complement(8642. .8949)
                     /rpt_family="AluSx"
     repeat_region   complement(8957. .8997)
                     /rpt_family="AT_rich"
     repeat_region   complement(9086. .9707)
                     /rpt_family="L2"
     repeat_region   complement(10244. .10409)
                     /rpt_family="L2"
     repeat_region   complement(10460. .10549)
                     /rpt_family="MIR"
     repeat_region   11002. .11191
                     /rpt_family="MIR"
     repeat_region   11198. .11379
                     /rpt_family="MER58A"
     repeat_region   11426. .11612
                     /rpt_family="MLT1A1"
     repeat_region   complement(11621. .11983)
                     /rpt_family="THE1C"
     repeat_region   11984. .12140
                     /rpt_family="MLT1A1"
     repeat_region   13004. .13306
                     /rpt_family="AluSp"
     repeat_region   complement(13644. .13765)
                     /rpt_family="MER45"
     repeat_region   complement(13795. .14093)
                     /rpt_family="AluSx"
     repeat_region   complement(14233. .14283)
                     /rpt_family="AT_rich"
     repeat_region   complement(14416. .14636)
                     /rpt_family="MIR"
     repeat_region   15412. .15712
                     /rpt_family="AluSg"
     repeat_region   15777. .15886
                     /rpt_family="MLT1E"
     repeat_region   16203. .16503
                     /rpt_family="AluY"
     repeat_region   16507. .16710
                     /rpt_family="MLT1E"
     repeat_region   complement(17148. .17445)
                     /rpt_family="AluSx"
     repeat_region   17836. .18475
                     /rpt_family="L2"
     repeat_region   20735. .21039
                     /rpt_family="AluSq"
     repeat_region   complement(21877. .21972)
                     /rpt_family="MER5A"
     repeat_region   22026. .22557
                     /rpt_family="MLT1F"
     repeat_region   complement(24081. .24160)
                     /rpt_family="(TAAA)n"
     repeat_region   complement(24678. .24977)
                     /rpt_family="AluSg"
     repeat_region   complement(25696. .26012)
                     /rpt_family="AluSx"
     repeat_region   27164. .27308
                     /rpt_family="MIR"
     repeat_region   complement(27491. .27587)
                     /rpt_family="L2"
     repeat_region   29508. .29808
                     /rpt_family="AluY"
     repeat_region   complement(30602. .30635)
                     /rpt_family="(TAAA)n"
     repeat_region   complement(30636. .30936)
                     /rpt_family="AluSx"
     repeat_region   31019. .31044
                     /rpt_family="AT_rich"
     repeat_region   31429. .31474
                     /rpt_family="AT_rich"
     repeat_region   32896. .33073
                     /rpt_family="L2"
     repeat_region   33368. .33517
                     /rpt_family="L2"
     repeat_region   complement(33894. .33934)
                     /rpt_family="AT_rich"
     unsure          34645. .34903
                     /note="double-stranded coverage, but plus strand read is
                     poor"
     unsure          34904. .35038
                     /note="only single-stranded coverage"
     repeat_region   complement(34999. .36185)
```

```
                     /rpt_family="L1MB7"
     repeat_region   36187. .36548
                     /rpt_family="THE1B"
     repeat_region   complement(36571. .39052)
                     /rpt_family="L1MB7"
     repeat_region   complement(39083. .39323)
                     /rpt_family="L1M4"
     repeat_region   40826. .41363
                     /rpt_family="MER41A"
     repeat_region   41398. .41522
                     /rpt_family="L2"
     repeat_region   complement(41535. .41602)
                     /rpt_family="L2"
     repeat_region   43073. .43112
                     /rpt_family="POLY_A"
     repeat_region   complement(44094. .44512)
                     /rpt_family="L1PA8"
     repeat_region   complement(44783. .45078)
                     /rpt_family="AluSg"
     repeat_region   45110. .45526
                     /rpt_family="MSTA"
     repeat_region   complement(45557. .45617)
                     /rpt_family="(GAAAA)n"
     repeat_region   complement(45618. .45919)
                     /rpt_family="AluJo"
     repeat_region   46148. .46261
                     /rpt_family="L2"
     repeat_region   complement(46948. .47255)
                     /rpt_family="MER4A"
     repeat_region   complement(47253. .47376)
                     /rpt_family="MER4A2"
     repeat_region   complement(47416. .47449)
                     /rpt_family="(CAAA)n"
     repeat_region   47816. .48059
                     /rpt_family="L2"
     repeat_region   48065. .48361
                     /rpt_family="AluSq"

  Query Match             0.2%;  Score 21;  DB 9;  Length 106793;
  Best Local Similarity   100.0%;  Pred. No. 83;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1356 TTTAAATGGAAAAACAAAAAA 1376
             |||||||||||||||||||||
Db      20723 TTTAAATGGAAAAACAAAAAA 20703


RESULT 74
AC096839_7
WPCOMMENT
  Sequence split into 8 fragments  LOCUS AC096839 Accession AC096839
     Fragment Name      Begin     End
     AC096839_0             1    110000
     AC096839_1        100001    210000
     AC096839_2        200001    310000
     AC096839_3        300001    410000
     AC096839_4        400001    510000
     AC096839_5        500001    610000
     AC096839_6        600001    710000
     AC096839_7        700001    809349
Continuation (8 of 8) of AC096839 from base 700001 (AC096839 Rattus norvegicus clone CH2

  Query Match             0.2%;  Score 21;  DB 2;  Length 109349;
  Best Local Similarity   100.0%;  Pred. No. 83;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2372 AAATAAATAAAAACAAGCAAA 2392
             |||||||||||||||||||||
Db      16800 AAATAAATAAAAACAAGCAAA 16820


RESULT 75
AC096839_2
WPCOMMENT
  Sequence split into 8 fragments  LOCUS AC096839 Accession AC096839
     Fragment Name      Begin     End
     AC096839_0             1    110000
     AC096839_1        100001    210000
     AC096839_2        200001    310000
     AC096839_3        300001    410000
     AC096839_4        400001    510000
     AC096839_5        500001    610000
     AC096839_6        600001    710000
     AC096839_7        700001    809349
Continuation (3 of 8) of AC096839 from base 200001 (AC096839 Rattus norvegicus clone

  Query Match             0.2%;  Score 21;  DB 2;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2372 AAATAAATAAAAACAAGCAAA 2392
             |||||||||||||||||||||
Db      38213 AAATAAATAAAAACAAGCAAA 38233


RESULT 76
AC096862_2/c
WPCOMMENT
  Sequence split into 5 fragments  LOCUS AC096862 Accession AC096862
     Fragment Name      Begin     End
     AC096862_0             1    110000
     AC096862_1        100001    210000
     AC096862_2        200001    310000
     AC096862_3        300001    410000
     AC096862_4        400001    450605
Continuation (3 of 5) of AC096862 from base 200001 (AC096862 Rattus norvegicus clone

  Query Match             0.2%;  Score 21;  DB 2;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6705 AAAAACAAAGATTTAAAAACA 6725
             |||||||||||||||||||||
Db      94389 AAAAACAAAGATTTAAAAACA 94369


RESULT 77
AC118188_1/c
WPCOMMENT
  Sequence split into 5 fragments  LOCUS AC118188 Accession AC118188
     Fragment Name      Begin     End
     AC118188_0             1    110000
     AC118188_1        100001    210000
     AC118188_2        200001    310000
     AC118188_3        300001    410000
     AC118188_4        400001    429365
Continuation (2 of 5) of AC118188 from base 100001 (AC118188 Rattus norvegicus clone

  Query Match             0.2%;  Score 21;  DB 2;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2372 AAATAAATAAAAACAAGCAAA 2392
             |||||||||||||||||||||
Db      23393 AAATAAATAAAAACAAGCAAA 23373


RESULT 78
AC122235_3
WPCOMMENT
  Sequence split into 5 fragments  LOCUS AC122235 Accession AC122235
     Fragment Name      Begin     End
     AC122235_0             1    110000
     AC122235_1        100001    210000
     AC122235_2        200001    310000
     AC122235_3        300001    410000
```

```
     AC122235_4          400001    434665
Continuation (4 of 5) of AC122235 from base 300001 (AC122235 Mus musculus chromosome UNK

  Query Match             0.2%;  Score 21;  DB 2;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4151 TGTTTTAAAAATGGAGAAAGA 4171
              |||||||||||||||||||||
Db      54032 TGTTTTAAAAATGGAGAAAGA 54052


RESULT 79
AC127191_2/c
WPCOMMENT
  Sequence split into 6 fragments  LOCUS AC127191 Accession AC127191
     Fragment Name      Begin     End
     AC127191_0             1    110000
     AC127191_1        100001    210000
     AC127191_2        200001    310000
     AC127191_3        300001    410000
     AC127191_4        400001    510000
     AC127191_5        500001    583874
Continuation (3 of 6) of AC127191 from base 200001 (AC127191 Rattus norvegicus clone CH2

  Query Match             0.2%;  Score 21;  DB 2;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2372 AAATAAATAAAAACAAGCAAA 2392
              |||||||||||||||||||||
Db      96139 AAATAAATAAAAACAAGCAAA 96119


RESULT 80
AC130200/c
LOCUS       AC130200              118998 bp    DNA     linear   HTG 11-JUN-2003
DEFINITION  Medicago truncatula clone mth1-64n13, WORKING DRAFT SEQUENCE.
ACCESSION   AC130200
VERSION     AC130200.25  GI:31581604
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 118998)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth1-64n13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 118998)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 118998)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Jun 11, 2003 this sequence version replaced gi:30231300.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   118998: contig of 118998 bp in length.
FEATURES            Location/Qualifiers
     source         1. .118998
                    /organism="Medicago truncatula"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:3880"
                    /clone="mth1-64n13"
                    /clone_lib="Medicago truncatula BAC library H1"
                    /note="This is one of two clones in the same well from
                    mth1-64n13"
BASE COUNT    41348 a  19508 c  18658 g  39484 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 2;  Length 118998;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2224 TATGTAGCAATATCAAGGGCT 2244
              |||||||||||||||||||||
Db      93949 TATGTAGCAATATCAAGGGCT 93929


RESULT 81
AR283580
LOCUS       AR283580              124884 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 76 from patent US 6528066.
ACCESSION   AR283580
VERSION     AR283580.1  GI:29720465
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 124884)
  AUTHORS   Grose,C.F. and Santos,R.
  TITLE     Variant varicella-zoster viruses and methods of use
  JOURNAL   Patent: US 6528066-A 76 04-MAR-2003;
FEATURES            Location/Qualifiers
     source         1. .124884
                    /organism="unknown"
BASE COUNT    33789 a  29295 c  28177 g  33623 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 6;  Length 124884;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        711 CATAGTATGCGTCGGTTCGCC 731
              |||||||||||||||||||||
Db      96832 CATAGTATGCGTCGGTTCGCC 96852


RESULT 82
BD097170
LOCUS       BD097170              124884 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Method for the quality control of attetunated varicella live
            vaccine.
ACCESSION   BD097170
VERSION     BD097170.1  GI:22642744
KEYWORDS    WO 0156600-A/1.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 124884)
  AUTHORS   Gomi,Y., Sunamachi,H., Takahashi,M. and Yamanishi,K.
  TITLE     Method for the quality control of attetunated varicella live
```

```
            vaccine
  JOURNAL   Patent: WO 0156600-A 1 09-AUG-2001;
            THE RESEARCH FOUNDATION FOR MICROBIAL DISEASES OF OSAKA UNIVERSITY,
            YASUYUKI GOMI,HIROKI SUNAMACHI,MICHIAKI TAKAHASHI,KOICHI YAMANISHI
COMMENT     OS   Varicella virus Dumas strain
            PN   WO 0156600-A/1
            PD   09-AUG-2001
            PF   31-JAN-2001 WO 2001JP000678
            PR   31-JAN-2000 JP 00P    62734
            PI   YASUYUKI GOMI,HIROKI SUNAMACHI,MICHIAKI TAKAHASHI,KOICHI  PI
            YAMANISHI
            PC   A61K39/25,A61P31/22,C12N15/38,C12N7/00,G01N33/15 CC   Method
            for the quality control of attetunated varicella live  CC
             vaccine
            FH   Key             Location/Qualifiers
            FT   source          1. .124884
            FT                   /organism='Varicella virus Dumas strain'.
FEATURES             Location/Qualifiers
     source          1. .124884
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT    33789 a  29295 c  28177 g  33623 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 6;  Length 124884;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        711 CATAGTATGCGTCGGTTCGCC 731
              |||||||||||||||||||||
Db      96832 CATAGTATGCGTCGGTTCGCC 96852


RESULT 83
HEVZVXX
LOCUS       HEVZVXX               124884 bp    DNA     linear   VRL 09-SEP-1999
DEFINITION  Varicella-Zoster virus complete genome.
ACCESSION   X04370 M14891 M16612
VERSION     X04370.1  GI:59989
KEYWORDS    capsid protein; deoxypyrimidine kinase; DNA polymerase; DNA-binding
            protein; dUTPase; exonuclease; genome; glycoprotein; kinase;
            nuclease; origin of replication; overlapping genes; polymerase;
            protein kinase; ribonucleotide reductase; thymidylate synthetase;
            unidentified reading frame.
SOURCE      Human herpesvirus 3
  ORGANISM  Human herpesvirus 3
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Alphaherpesvirinae; Varicellovirus.
REFERENCE   1  (bases 1 to 124884)
  AUTHORS   Davison,A.J. and Scott,J.E.
  TITLE     The complete DNA sequence of varicella-zoster virus
  JOURNAL   J. Gen. Virol. 67 (Pt 9), 1759-1816 (1986)
  MEDLINE   86306657
   PUBMED   3018124
COMMENT     Data kindly reviewed (07-SEP-1987) by Davison A.J.
FEATURES             Location/Qualifiers
     source          1. .124884
                     /organism="Human herpesvirus 3"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10335"
     misc_feature    88. .89
                     /note="TRL/UL boundary"
     misc_feature    complement(588. .593)
                     /note="polyA signal (ORF1)"
     CDS             complement(589. .915)
                     /note="ORF1 (AA1-108)"
                     /codon_start=1
                     /protein_id="CAA27884.1"
                     /db_xref="GI:59990"
                     /db_xref="SWISS-PROT:P09266"
                     /translation="MSRVSEYGVPEGVRESDSDTDSVFMYQHTELMQNNASPLVVQTR
                     PPAVLIPLVDVPRPRSRRKASAQLKMQMDRLCNVLGVVLQMATLALVTYIAFVVHTRA
                     TSCKRE"
     CDS             1134. .1850
                     /note="ORF2 (AA1-238)"
                     /codon_start=1
                     /protein_id="CAA27885.1"
                     /db_xref="GI:59991"
                     /db_xref="SWISS-PROT:P09267"
                     /translation="MHVISETLAYGHVPAFIMGSTLVRPSLNATAEENPASETRCLLR
                     VLAGRTVDLPGGGTLHITCTKTYVIIGKYSKPGERLSLARLIGRAMTPGGARTFIILA
                     MKEKRSTTLGYECGTGLHLLAPSMGTFLRTHGLSNRDLCLWRGNIYDMHMQRLMFWEN
                     IAQNTTETPCITSTLTCNLTEDSGEAALTTSDRPTLPTLTAQGRPTVSNIRGILKGSP
                     RQQPVCHRVRFAEPTEGVLM"
     misc_feature    1857. .1862
                     /note="polyA signal (ORF2)"
     misc_feature    complement(1890. .1895)
                     /note="polyA signal (ORF3)"
     CDS             complement(1908. .2447)
                     /note="ORF3 (AA1-179)"
                     /codon_start=1
                     /protein_id="CAA27886.1"
                     /db_xref="GI:59992"
                     /db_xref="SWISS-PROT:P09268"
                     /translation="MDTTGASESSQPIRVNLKPDPLASFTQVIPPLALETTWTCPANS
                     HAPTPSPLYGVKRLCALRATCGRADDLHAFLIGLGRRDKPSESPMYVDLQPFCSLLNS
                     QRLLPEMANYNTLCDAPFSAATQQMMLESGQLGVHLAAIGYHCHCKSPFSAECWTGAS
                     EAYDHVVCGGKARAAVGGL"
     misc_feature    complement(2777. .2782)
                     /note="polyA signal (ORF4)"
     CDS             complement(2783. .4141)
                     /note="ORF4 (AA1-452)"
                     /codon_start=1
                     /protein_id="CAA27887.1"
                     /db_xref="GI:59993"
                     /db_xref="SWISS-PROT:P09269"
                     /translation="MASASIPTDPDVSTICEDFMNLLPDEPSDDFALEVTDWANDEAI
                     GSTPGEDSTTSRTVYVERTADTAYNPRYSKRRHGRRESYHHNRPKTLVVVLPDSNHHG
                     GRDVETGYARIERGHRRSSRSYNTQSSRKHRDRSLSNRRRRPTTPPAMTTGERNDQTH
                     DESYRLRFSKRDARRERIRKEYDIPVDRITGRAIEVVSTAGASVTIDSVRHLDETIEK
                     LVVRYATIQEGDSWASGGCFPGIKQNTSWPELMLYGHELYRTFESYKMDSRIARALRE
                     RVIRGESLIEALESADELLTWIKMLAAKNLPIYTNNPIVATSKSLLENLKLKLGPFVR
                     CLLLNRDNDLGSRTLPELLRQQRFSDITCITTYMFVMIARIANIVVRGSKFVEYDDIS
                     CNVQVLQEYTPGSCLAGVLEALITHQRECGRVECTLSTWAGHLSDARPYGKYFKCSTF
                     NC"
     CDS             complement(4252. .5274)
                     /note="ORF5 (AA1-340)"
                     /codon_start=1
                     /protein_id="CAA27888.1"
                     /db_xref="GI:59994"
                     /db_xref="SWISS-PROT:P09261"
                     /translation="MQALGIKTEHFIIMCLLSGHAVFTLWYTARVKFEHECVYATTVI
                     NGGPVVWGSYNNSLIYVTFVNHSTFLDGLSGYDYSCRENLLSGDTMVKTAISTPLHDK
                     IRIVLGTRNCHAYFWCVQLKMIFFAWFVYGMYLQFRRIRRMFGPFRSSCELISPTSYS
                     LNYVTRVISNILLGYPYTKLARLLCDVSMRRDGMSKVFNADPISFLYMHKGVTLLMLL
                     EVIAHISSGCIVLLTLGVAYTPCALLYPTYIRILAWVVVCTLAIVELISYVRPKPTKD
                     NHLNHINTGGIRGICTTCCATVMSGLAIKCFYIVIFAIAVVIFMHYEQRVQVSLFGES
                     ENSQKH"
     CDS             complement(5326. .8577)
                     /note="ORF6 (AA1-1083)"
                     /codon_start=1
                     /protein_id="CAA27889.1"
                     /db_xref="GI:59995"
                     /db_xref="SWISS-PROT:P09270"
                     /translation="MDKSSKPTIRLLFATKGCAISHSLLLLTGQISTEPLYVVSYTWT
                     PDLDDVFVKNGREEITQVIPTKRPREVTENDEENQIMHLFCSRDVNVIFYLIGGFSTG
                     DVRSRVWPIFFCCFKTQTDFKALYKALWYGAPLNPHIISDTLCISETFDIHSEVIQTL
                     MVTTHHLNRKGLSDNGLCITEATLCKLVKKSVGRQELTSLYAHYERQVLAAYRRLYWG
                     YGCSPFWYIVRFGPSEKTLVLATRYYLLQTDTSYNTLETPLYDLQAIKDLFLTYQVPA
                     LPNCSGYNISDLLSFDKLSMFCCSSTYTRGLTAKNALSYILQRIHTDTTEIHAVSEYI
                     TNDRKGLKVPDREFVDYIYLAHFECFNRKQIADHLQAVTYSDFVNKPVLLKSSNLGKR
                     ATANFFNHVRSRLNMRDYIKKNVICDVTELGPEIGHKYTITKTYTLSLTYAAKPSKFI
                     GVCDLATTLTRRVENIEKQFSPYGWSSTIPSNPPGFDELSNFEDSGVSAEALRAANFA
                     NDTPNQSGRTGFDTSPGITKLLLFFSAATGIATHDVSILSYKTPLEALIGHSEVTGPM
                     PVYRVALPHGAQAFAVIANDTWSSITNRYTLPHEARLIAEDLKQINPCNFVAASLRDM
```

```
                     QLTLLLSTSVKNVSKISSNIPKDQLYINRNELFNTNLIITNLILDVDFHIRKPIPLGI
                     LHAGMRAFRHGILTAMQLLFPKAVVNPNKDPCYFYKTACPEPTVEVLDDDNLLDITSH
                     SDIDFYIENGELYTCVEENYTEDVWFFDTQTTSEVHTHADVSNNENLHETLPCNCKEK
                     IGFRVCVPIPNPYALVGSSTLKGFAQILQQAVLLEREFVEYIGPYLRDFSFIDTGVYS
                     HGHSLRLPFFSKVTTTGTAVGQLLPFYVVPEQCIDILAFVTSHRNPANFHFHSRPQSN
                     VPVQFILHNLGGEYAEFFERKVARNKQIFSSPQISLTKALKERGVTCLDAFTLEAFVD
                     STILESIVEHIAVHFPGRDREYTLTSSKCIAIKRDWVLFQLICGTKGFTCLRYPHRGG
                     RTAPRTFVSLRVDHHNRLCISLAQQCFATKCDSNRMHTIFTLEVPNYPNLTSS"
CDS                  8607. .9386
                     /note="ORF7 (AA1-259)"
                     /codon_start=1
                     /protein_id="CAA27890.1"
                     /db_xref="GI:59996"
                     /db_xref="SWISS-PROT:P09271"
                     /translation="MQTVCASLCGYARIPTEEPSYEEVRVNTHPQGAALLRLQEALTA
                     VNGLLPAPLTLEDVVASADNTRRLVRAQALARTYAACSRNIECLKQHHFTEDNPGLNA
                     VVRSHMENSKRLADMCLAAITHLYLSVGAVDVTTDDIVDQTLRMTAESEVVMSDVVLL
                     EKTLGVVAKPQASFDVSHNHELSIAKGENVGLKTSPIKSEATQLSEIKPPLIEVSDNN
                     TSNLTKKTYPTETLQPVLTPKQTQDVQRTTPAIKKSHVMLV"
misc_feature         9394. .9399
                     /note="polyA signal (ORF7)"
misc_feature         complement(9426. .9431)
                     /note="polyA signal (ORF8)"
CDS                  complement(9477. .10667)
                     /note="ORF8 (AA1-396)"
                     /codon_start=1
                     /protein_id="CAA27891.1"
                     /db_xref="GI:59997"
                     /db_xref="SWISS-PROT:P09254"
                     /translation="MNEAVIDPILETAVNTGDMFCSQTIPNRCLKDTILIEVQPECAD
                     TLQCVLDDKVSRHQPLLLRNHKKLELPSEKSVTRGGFYMQQLELLVKSAPPNEYALLL
                     IQCKDTALADEDNFFVANGVIDAGYRGVISALLYYRPGVTVILPGHLTIYLFPVKLRQ
                     SRLLPKNVLKHLDPIFKSIQVQPLSNSPSNYEKPVIPEFADISTVQQGQPLHRDSAEY
                     HIDVPLTYKHIINPKRQEDAGYDICVPYNLYLKRNEFIKIVLPIIRDWDLQHPSINAY
                     IFGRSSKSRSGIIVCPTAWPAGEHCKFYVYNLTGDDIRIKTGDRLAQVLLIDHNTQIH
                     LKHNVLSNIAFPYAIRGKCGIPGVQWYFTKTLDLIATPSERGTRGFGSTDKETNDVDF
                     LLKH"
CDS                  11009. .11917
                     /note="ORF9 (AA1-302)"
                     /codon_start=1
                     /protein_id="CAA27892.1"
                     /db_xref="GI:59998"
                     /db_xref="SWISS-PROT:P09272"
                     /translation="MASSDGDRLCRSNAVRRKTTPSYSGQYRTARRSVVVGPPDDSDD
                     SLGYITTVGADSPSPVYADLYFEHKNTTPRVHQPNDSSGSEDDFEDIDEVVAAFREAR
                     LRHELVEDAVYENPLSVEKPSRSFTKNAAVKPKLEDSPKRAPPGAGAIASGRPISFST
                     APKTATSSWCGPTPSYNKRVFCEAVRRVAAMQAQKAAEAAWNSNPPRNNAELDRLLTG
                     AVIRITVHEGLNLIQAANEADLGEGASVSKRGHNRKTGDLQGGMGNEPMYAQVRKPKS
                     RTDTQTTGRITNRSRARSASRTDTRK"
misc_feature         11959. .11964
                     /note="polyA signal (ORF9)"
CDS                  12160. .13392
                     /note="ORF10 (AA1-410)"
                     /codon_start=1
                     /protein_id="CAA27893.1"
                     /db_xref="GI:59999"
                     /db_xref="SWISS-PROT:P09265"
                     /translation="MECNLGTEHPSTDTWNRSKTEQAVVDAFDESLFGDVASDIGFET
                     SLYSHAVKTAPSPPWVASPKILYQQLIRDLDFSEGPRLLSCLETWNEDLFSCFPINED
                     LYSDMMVLSPDPDDVISTVSTKDHVEMFNLTTRGSVRLPSPPKQPTGLPAYVQEVQDS
                     FTVELRAREEAYTKLLVTYCKSIIRYLQGTAKRTTIGLNIQNPDQKAYTQLRQSILLR
                     YYREVASLARLLYLHLYLTVTREFSWRLYASQSAHPDVFAALKFTWTERRQFTCAFHP
                     VLCNHGIVLLEGKPLTASALREINYRRRELGLPLVRCGLVEENKSPLVQQPSFSVHLP
                     RSVGFLTHHIKRKLDAYAVKHPQEPRHVRADHPYAKVVENRNYGSSIEAMILAPPSPS
                     EILPGDPPRPPTCGFLTR"
misc_feature         13416. .13421
                     /note="polyA signal (ORF10)"
CDS                  13590. .16049
                     /note="ORF11 (AA1-819)"
                     /codon_start=1
                     /protein_id="CAA27894.1"
                     /db_xref="GI:60000"
                     /db_xref="SWISS-PROT:P09263"
                     /translation="MQSGHYNRRQSRRQRISSNTTDSPRHTHGTRYRSTNWYTHPPQI
                     LSNSETLVAVQELLNSEMDQDSSSDASDDFPGYALHHSTYNGSEQNTSTSRHENRIFK
                     LTEREANEEININTDAIDDEGEAEEGEAEEDAIDDEGEAEEGEAEEDAIDDEGEAEEG
                     EAEEDAIDDEGEAEEGEAEEGEAEEDAIDDEGEAEEDAAEEDAIDDEGEAEED
                     YFSVSQVCSRDADEVYFTLDPEISYSTDLRIAKVMEPAVSKELNVSKRCVEPVTLTGS
                     MLAHNGFDESWFAMRECTRREYITVQGLYDPIHLRYQFDTSRMTPPQILRTIPALPNM
                     TLGELLLIFPIEFMAQPISIERILVEDVFLDRRASSKTHKYGPRWNSVYALPYNAGKM
                     YVQHIPGFYDVSLRAVGQGTAIWHHMILSTAACAISNRISHGDGLGFLLDAAIRISAN
                     CIFLGRNDNFGVGDPCWLEDHLAGLPREAVPDVLQVTQLVLPNRGPTVAIMRGFFGAL
                     AYWPELRIAISEPSTSLVRYATGHMELAEWFLFSRTHSLKPQFTPTEREMLASFFTLY
                     VTLGGGMLNWICRATAMYLAAPYHSRSAYIAVCESLPYYYIPVNSDLLCDLEVLLLGE
                     VDLPTVCESYATIAHELTGYEAVRTAATNFMIEFADCYKESETDLMVSAYLGAVLLLQ
                     RVLGHANLLLLLLSGAALYGGCSIYIPRGILDAYNTLMLAASPLYAHQTLTSFWKDRD
                     DAMQTLGIRPTTDVLPKEQDRIVQASPIEMNFRFVGLETIYPREQPIPSVDLAENLMQ
                     YRNEILGLDWKSVAMHLLRKY"
repeat_region        13937. .14242
                     /note="reiteration R1"
```

```
  Query Match             0.2%;  Score 21;  DB 14;  Length 124884;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        711 CATAGTATGCGTCGGTTCGCC 731
              |||||||||||||||||||||
Db      96832 CATAGTATGCGTCGGTTCGCC 96852
```

```
RESULT 84
AB097932
LOCUS       AB097932              125078 bp    DNA     linear   VRL 13-DEC-2002
DEFINITION  Human herpesvirus 3 ORF62 gene for IE62 transactivator, complete
            cds, strain: The Oka vaccine.
ACCESSION   AB097932
VERSION     AB097932.1  GI:26665420
KEYWORDS    .
SOURCE      Human herpesvirus 3
  ORGANISM  Human herpesvirus 3
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Alphaherpesvirinae; Varicellovirus. .
REFERENCE   1
  AUTHORS   Gomi,Y., Sunamachi,H., Mori,Y., Nagaike,K., Takahashi,M. and
            Yamanishi,K.
  TITLE     Comparison of the complete DNA sequences of the Oka varicella
            vaccine and its parental virus
  JOURNAL   J. Virol. 76 (22), 11447-11459 (2002)
  MEDLINE   22276345
   PUBMED   12388706
REFERENCE   2  (bases 1 to 125078)
  AUTHORS   Gomi,Y., Sunamachi,H., Nagaike,K., Mori,Y., Takahashi,M. and
            Yamanishi,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2002) Yasuyuki Gomi, Kanonji Institute, the
            research foundation for microbial diseases of Osaka university,
            research division; 2-9-41 Yahata-cho, Kanonji-city, Kagawa
            768-0062, Japan (E-mail:ygomi@mail.biken.or.jp,
            Tel:81-875-25-4171(ex.620), Fax:81-875-25-4843)
FEATURES             Location/Qualifiers
     source          1. .125078
                     /organism="Human herpesvirus 3"
                     /mol_type="genomic DNA"
                     /strain="The Oka vaccine"
                     /db_xref="taxon:10335"
                     /note="synonym: varicella-zoster virus"
     gene            complement(105189. .109121)
                     /gene="ORF62"
     CDS             complement(105189. .109121)
                     /gene="ORF62"
                     /codon_start=1
                     /product="IE62 transactivator"
                     /protein_id="BAC44881.1"
                     /db_xref="GI:26665421"
                     /translation="MDTPPMQRSTPQRAGSPDTLELMDLLDAAAAAAEHRARVVTSSQ
                     PDDLLFGENGVMVGREHEIVSIPSVSGLQPEPRTEDVGEELTQDDYVCEDGQDLXGSP
```

```
                     VIPLAEVFHTRFSEAGAREPTGADRSLETVSLGTKLARSPKPPMNDGETGRGTTPPFP
                     QAFSPVSPASPVGDAAGNDQREDQRSIPRQTTRGNSPGLPSVVHRDRQTQSISGKKPG
                     DEQAGHAHASGDGVVLQKTQRPAQGKSPKKKTLKVKVPLPARKPGGPVPGPVEQLYHV
                     LSDSVPAKGAKADLPFETDDTRPRKHDARGITPRVPGRSSGGKPRAFLALPGRSHAPD
                     PIEDDSPVEKKPKSREFVSSSSSSSSWGSSSEDEDDEPRRVSVGSETTGSRSGREHAP
                     SPSNSDDSDSNDGGSTKQNIQPGYRSISGPDPRIRKTKRLAGEPGRQRQKSFSLPRSR
                     TPIIPPVSGPLMMPDGSPWPGSAPLPSNRVRFGPSGETREGHWEDEAVRAARARYEAS
                     TEPXPLYVPELGDPARQYRALINLIYCPDRDPIAWLQNPKLTGVNSALNQFYQKLLPP
                     GRAGTAVTGSVASPVPHVGEAMATGEALWALPHAAAAVAMSRRYDRAQKHFILQSLRR
                     AFAGMAYPEATGSSPAARISRGHPSPTTPATQTPDPQPSAAARSLSVCPPDDRLRTPR
                     KRKSQPVESRSLLDKIRETPVADARVADDHVVSKAKRRVSEPVTITSGPVVDPPAVIT
                     MPLDGPAPNGGFRRIPRGALHTPVPSDQARKAYCTPETIARLVDDPLFPTAWRPALSF
                     DPGALAEIAARRPGGGDRRFGPPSGVEALRRRCAWMRQIPDPEDVRLLIIYDPLPGED
                     INGPLESTLATDPGPSWSPSRGGLSVVLAALSNRLCLPSTHAWAGNWTGPPDVSALNA
                     RGVLLLSTRDLAFAGAVEYLGSRLASARRRLLVLDAVALERWPGDGPALSQYHVYVRA
                     PARPDAQAVVRWPDSAVTEGLARAVFASSRTFGPASFARIETAFANLYPGEQPLCLCR
                     GGNVAYTVCTRAGPKTRVPLSPREYRQYVLPGFDGCKDLARQSRGLGLGAADFVDEAA
                     HSHRAANRWGLGAALRPVFLPEGRRPGAAGPEAGDVPTWARVFCRHALLEPDPAAEPL
                     VLPPVAGRSVALYASADEARNALPPIPRVMWPPGFGAAETVLEGSDGTRFAFGHHGGS
                     ERPAETQAGRQRRTADDREHALEPDDWEVGCEDAWDSEEGGGDDGDAPGSSFGVSVVS
                     VAPGVLRDRRVGXRPAVKVELLSSSSSSEDEDDVWGGRGGRSPPQSRG"
BASE COUNT    33768 a  29398 c  28263 g  33619 t     30 others
ORIGIN

  Query Match             0.2%;  Score 21;  DB 14;  Length 125078;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        711 CATAGTATGCGTCGGTTCGCC 731
              |||||||||||||||||||||
Db      96706 CATAGTATGCGTCGGTTCGCC 96726


RESULT 85
AB097933
LOCUS       AB097933               125125 bp    DNA     linear   VRL 13-DEC-2002
DEFINITION  Human herpesvirus 3 ORF62 gene for IE62 transactivator, complete
            cds, strain: The Oka parental.
ACCESSION   AB097933
VERSION     AB097933.1  GI:26665422
KEYWORDS    .
SOURCE      Human herpesvirus 3
  ORGANISM  Human herpesvirus 3
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Alphaherpesvirinae; Varicellovirus.
REFERENCE   1
  AUTHORS   Gomi,Y., Sunamachi,H., Mori,Y., Nagaike,K., Takahashi,M. and
            Yamanishi,K.
  TITLE     Comparison of the complete DNA sequences of the Oka varicella
            vaccine and its parental virus
  JOURNAL   J. Virol. 76 (22), 11447-11459 (2002)
  MEDLINE   22276345
   PUBMED   12388706
REFERENCE   2  (bases 1 to 125125)
  AUTHORS   Gomi,Y., Sunamachi,H., Nagaike,K., Mori,Y., Takahashi,M. and
            Yamanishi,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2002) Yasuyuki Gomi, Kanonji institute, the
            research foundation for microbial diseases of Osaka university,
            research division; 2-9-41 Yahata-cho, Kanonji-city, Kagawa
            768-0062, Japan (E-mail:ygomi@mail.biken.or.jp,
            Tel:81-875-25-4171(ex.620), Fax:81-875-25-4843)
FEATURES             Location/Qualifiers
     source          1. .125125
                     /organism="Human herpesvirus 3"
                     /mol_type="genomic DNA"
                     /strain="The Oka parental"
                     /db_xref="taxon:10335"
                     /note="synonym: varicella-zoster virus"
     gene            complement(105166. .109098)
                     /gene="ORF62"
     CDS             complement(105166. .109098)
                     /gene="ORF62"
                     /codon_start=1
                     /product="IE62 transactivator"
                     /protein_id="BAC44882.1"
                     /db_xref="GI:26665423"
                     /translation="MDTPPMQRSTPQRAGSPDTLELMDLLDAAAAAAEHRARVVTSSQ
                     PDDLLFGENGVMVGREHEIVSIPSVSGLQPEPRTEDVGEELTQDDYVCEDGQDLMGSP
                     VIPLAEVFHTRFSEAGAREPTGADRSLETVSLGTKLARSPKPPMNDGETGRGTTPPFP
                     QAFSPVSPASPVGDAAGNDQREDQRSIPRQTTRGNSPGLPSVVHRDRQTQSISGKKPG
                     DEQAGHAHASGDGVVLQKTQRPAQGKSPKKKTLKVKVPLPARKPGGPVPGPVEQLYHV
                     LSDSVPAKGAKADLPFETDDTRPRKHDARGITPRVPGRSSGGKPRAFLALPGRSHAPD
                     PIEDDSPVEKKPKSREFVSSSSSSSSWGSSSEDEDDEPRRVSVGSETTGSRSGREHAP
                     SPSNSDDSDSNDGGSTKQNIQPGYRSISGPDPRIRKTKRLAGEPGRQRQKSFSLPRSR
                     TPIIPPVSGPLMMPDGSPWPGSAPLPSNRVRFGPSGETREGHWEDEAVRAARARYEAS
                     TEPVPLYVPELGDPARQYRALINLIYCPDRDPIAWLQNPKLTGVNSALNQFYQKLLPP
                     GRAGTAVTGSVASPVPHVGEAMATGEALWALPHAAAAVAMSRRYDRAQKHFILQSLRR
                     AFASMAYPEATGSSPAARISRGHPSPTTPATQTPDPQPSAAARSLSVCPPDDRLRTPR
                     KRKSQPVESRSLLDKIRETPVADARVADDHVVSKAKRRVSEPVTITSGPVVDPPAVIT
                     MPLDGPAPNGGFRRIPRGALHTPVPSDQARKAYCTPETIARLVDDPLFPTAWRPALSF
                     DPGALAEIAARRPGGGDRRFGPPSGVEALRRRCAWMRQIPDPEDVRLLIIYDPLPGED
                     INGPLESTLATDPGPSWSPSRGGLSVVLAALSNRLCLPSTHAWAGNWTGPPDVSALNA
                     RGVLLLSTRDLAFAGAVEYLGSRLASARRRLLVLDAVALERWPRDGPALSQYHVYVRA
                     PARPDAQAVVRWPDSAVTEGLARAVFASSRTFGPASFARIETAFANLYPGEQPLCLCR
                     GGNVAYTVCTRAGPKTRVPLSPREYRQYVLPGFDGCKDLARQSRGLGLGAADFVDEAA
                     HSHRAANRWGLGAALRPVFLPEGRRPGAAGPEAGDVPTWARVFCRHALLEPDPAAEPL
                     VLPPVAGRSVALYASADEARNALPPIPRVMWPPGFGAAETVLEGSDGTRFVFGHHGGS
                     ERPAETQAGRQRRTADDREHALEPDDWEVGCEDAWDSEEGGGDDGDAPGSSFGVSIVS
                     VAPGVLRDRRVGLRPAVKVELLSSSSSSEDEDDVWGGRGGRSPPQSRG"
BASE COUNT    33796 a  29413 c  28260 g  33656 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 14;  Length 125125;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        711 CATAGTATGCGTCGGTTCGCC 731
              |||||||||||||||||||||
Db      96682 CATAGTATGCGTCGGTTCGCC 96702


RESULT 86
BD097171
LOCUS       BD097171               125157 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Method for the quality control of attetunated varicella live
            vaccine.
ACCESSION   BD097171
VERSION     BD097171.1  GI:22642745
KEYWORDS    WO 0156600-A/2.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 125157)
  AUTHORS   Gomi,Y., Sunamachi,H., Takahashi,M. and Yamanishi,K.
  TITLE     Method for the quality control of attetunated varicella live
            vaccine
  JOURNAL   Patent: WO 0156600-A 2 09-AUG-2001;
            THE RESEARCH FOUNDATION FOR MICROBIAL DISEASES OF OSAKA UNIVERSITY,
            YASUYUKI GOMI,HIROKI SUNAMACHI,MICHIAKI TAKAHASHI,KOICHI YAMANISHI
COMMENT     OS   Attenuated varicella virus Oka strain
            PN   WO 0156600-A/2
            PD   09-AUG-2001
            PF   31-JAN-2001 WO 2001JP000678
            PR   31-JAN-2000 JP 00P    62734
            PI   YASUYUKI GOMI,HIROKI SUNAMACHI,MICHIAKI TAKAHASHI,KOICHI  PI
            YAMANISHI
            PC   A61K39/25,A61P31/22,C12N15/38,C12N7/00,G01N33/15 CC   Method
            for the quality control of attetunated varicella live  CC
             vaccine
            FH   Key             Location/Qualifiers
            FT   source          1. .125157
            FT                   /organism='Attenuated varicella virus Oka  FT
                                 strain'.
FEATURES             Location/Qualifiers
     source          1. .125157
```

```
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT    33776 a   29439 c   28285 g   33622 t      35 others
ORIGIN

  Query Match             0.2%;  Score 21;  DB 6;  Length 125157;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        711 CATAGTATGCGTCGGTTCGCC 731
              |||||||||||||||||||||
Db      96778 CATAGTATGCGTCGGTTCGCC 96798


RESULT 87
AC139626/c
LOCUS       AC139626               128091 bp    DNA     linear   HTG 06-MAR-2003
DEFINITION  Takifugu rubripes clone 196F23, WORKING DRAFT SEQUENCE, 9 ordered
            pieces.
ACCESSION   AC139626
VERSION     AC139626.2  GI:28866984
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Takifugu rubripes (Fugu rubripes)
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 128091)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Cariaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
            Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J.,
            Paguirigan,C., Pearson,R., Portnoy,M.E., Prasad,A.,
            Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
            Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 128091)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 128091)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Mar 6, 2003 this sequence version replaced gi:28269376.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: ego
                 Center clone name: 196F23

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 126859 bases at least Q40
                 Consensus quality: 127065 bases at least Q30
                 Consensus quality: 127213 bases at least Q20
                 Insert size: 128000; agarose-fp
                 Insert size: 127291; sum-of-contigs
                 Quality coverage: 10.76x in Q20 bases; agarose-fp
                 Quality coverage: 10.82x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    23378: contig of 23378 bp in length
            *    23379    23478: gap of unknown length
            *    23479    29206: contig of 5728 bp in length
            *    29207    29306: gap of unknown length
            *    29307    41054: contig of 11748 bp in length
            *    41055    41154: gap of unknown length
            *    41155    45491: contig of 4337 bp in length
            *    45492    45591: gap of unknown length
            *    45592    84142: contig of 38551 bp in length
            *    84143    84242: gap of unknown length
            *    84243    92572: contig of 8330 bp in length
            *    92573    92672: gap of unknown length
            *    92673    95251: contig of 2579 bp in length
            *    95252    95351: gap of unknown length
            *    95352   118176: contig of 22825 bp in length
            *   118177   118276: gap of unknown length
            *   118277   128091: contig of 9815 bp in length.
FEATURES             Location/Qualifiers
     source          1. .128091
                     /organism="Takifugu rubripes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:31033"
                     /clone="196F23"
                     /clone_lib="Incyte Genomics"
     misc_feature    1. .23378
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    23479. .29206
                     /note="assembly_fragment"
     misc_feature    29307. .41054
                     /note="assembly_fragment"
     misc_feature    41155. .45491
                     /note="assembly_fragment"
     misc_feature    45592. .84142
                     /note="assembly_fragment"
     misc_feature    84243. .92572
                     /note="assembly_fragment"
     misc_feature    92673. .95251
                     /note="assembly_fragment"
     misc_feature    95352. .118176
                     /note="assembly_fragment"
     misc_feature    118277. .128091
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    35675 a   28483 c   28261 g   34872 t     800 others
ORIGIN

  Query Match             0.2%;  Score 21;  DB 2;  Length 128091;
  Best Local Similarity   100.0%;  Pred. No. 84;
```

```
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4234 TTCTGGTGACATTTTACTCCA 4254
              ||||||||||||||||||||
Db      31691 TTCTGGTGACATTTTACTCCA 31671


RESULT 88
AC130257
LOCUS       AC130257              129016 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-414I4, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC130257
VERSION     AC130257.4  GI:25085209
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 129016)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 129016)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 129016)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:23907812.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GYOZ
                Center clone name: CH230-414I4
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 118746 bases at least Q40
                Consensus quality: 120078 bases at least Q30
                Consensus quality: 121002 bases at least Q20
                Estimated insert size: 120140; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   127610: contig of 127610 bp in length
            *   127611   127710: gap of unknown length
            *   127711   129016: contig of 1306 bp in length.
FEATURES             Location/Qualifiers
     source          1. .129016
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-414I4"
     misc_feature    1109. .52336
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BZ169779"
     misc_feature    complement(117097. .118032)
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BZ169777"
     misc_feature    119665. .120701
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    126431. .127610
```

```
                     /note="wgs_end_extension
                     clone_end:T7"
BASE COUNT    31872 a  27474 c  28258 g  33962 t   7450 others
ORIGIN

  Query Match             0.2%;  Score 21;  DB 2;  Length 129016;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1967 CACTCAGTGATGACCTCAGGA 1987
              |||||||||||||||||||||
Db      56215 CACTCAGTGATGACCTCAGGA 56235


RESULT 89
AL808113
LOCUS       AL808113              129039 bp    DNA     linear   ROD 07-SEP-2002
DEFINITION  Mouse DNA sequence from clone RP23-464N8 on chromosome X, complete
            sequence.
ACCESSION   AL808113
VERSION     AL808113.7  GI:22798407
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 129039)
  AUTHORS   Howden,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 12, 2002 this sequence version replaced gi:22204742.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-464N8 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .129039
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="X"
                     /clone="RP23-464N8"
                     /clone_lib="RPCI-23"
BASE COUNT    43218 a  24811 c  24616 g  36394 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 10;  Length 129039;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2545 GTCTCTTAGGGCCCTGGGTTG 2565
              |||||||||||||||||||||
Db      82589 GTCTCTTAGGGCCCTGGGTTG 82609


RESULT 90
AL354701
LOCUS       AL354701              131366 bp    DNA     linear   PRI 16-OCT-2001
DEFINITION  Human DNA sequence from clone RP11-785H23 on chromosome 9, complete
            sequence.
ACCESSION   AL354701
VERSION     AL354701.7  GI:16214576
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 131366)
  AUTHORS   Ramsay,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 17, 2001 this sequence version replaced gi:15041814.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-785H23 is from the library RPCI-11.3 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-785H23 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP11-138L24 is at 129367 in this
            sequence. The true right end of clone RP11-342F21 is at 2000 in
            this sequence.
FEATURES             Location/Qualifiers
     source          1. .131366
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-785H23"
                     /clone_lib="RPCI-11.3"
     misc_feature    complement(22450. .22985)
                     /note="match: STS: Em:G56192"
     misc_feature    complement(59216. .59581)
                     /note="match: STS: Em:G51747"
BASE COUNT    41024 a  22825 c  22986 g  44531 t
ORIGIN
```

```
  Query Match             0.2%;  Score 21;  DB 9;  Length 131366;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3921 CTATCTCTTTAAAAATATTTT 3941
              |||||||||||||||||||||
Db      44358 CTATCTCTTTAAAAATATTTT 44378


RESULT 91
AL929509
LOCUS       AL929509              139407 bp    DNA     linear   HTG 10-JUN-2003
DEFINITION  Danio rerio clone CH211-165I22, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AL929509
VERSION     AL929509.11  GI:31616703
KEYWORDS    HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 139407)
  AUTHORS   Sycamore,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 11, 2003 this sequence version replaced gi:31559223.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC165I22
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 139296 bases at least Q40
            Consensus quality: 139391 bases at least Q30
            Consensus quality: 139405 bases at least Q20
            Insert size: 139407; sum-of-contigs
            Insert size: 141700; 2.6% error; agarose-fp
            Quality coverage: 7.54x in Q20 bases; sum-of-contigs Quality
            coverage: 7.42x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .139407
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-165I22"
                     /clone_lib="CHORI-211"
     misc_feature    1. .139407
                     /note="assembly_fragment:00072
                     clone_end:T7
                     vector_side:left"
BASE COUNT    43592 a  25580 c  25480 g  44755 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 2;  Length 139407;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4843 TATATTTTAAGGCATAATGCA 4863
              |||||||||||||||||||||
Db     104453 TATATTTTAAGGCATAATGCA 104473


RESULT 92
AP004124/c
LOCUS       AP004124              140488 bp    DNA     linear   HTG 21-MAR-2002
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 2 clone
            OJ1669_F01, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AP004124
VERSION     AP004124.1  GI:15375110
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, BAC
            clone:OJ1669_F01
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 140488)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2001) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The nucleotide sequence of this BAC clone was generated by
            combining Monsanto and RGP-Japan sequencing data.
            NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .140488
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="2"
                     /clone="OJ1669_F01"
BASE COUNT    38703 a  32049 c  32082 g  37553 t    101 others
ORIGIN

  Query Match             0.2%;  Score 21;  DB 2;  Length 140488;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1964 TGCCACTCAGTGATGACCTCA 1984
              |||||||||||||||||||||
Db      99391 TGCCACTCAGTGATGACCTCA 99371


RESULT 93
AL590678/c
LOCUS       AL590678              149849 bp    DNA     linear   PRI 09-MAY-2001
DEFINITION  Human DNA sequence from clone RP11-141E7 on chromosome 6, complete
            sequence.
ACCESSION   AL590678
VERSION     AL590678.3  GI:14018314
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 149849)
  AUTHORS   Williams,S.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 11, 2001 this sequence version replaced gi:13751625.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-141E7 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-141E7 The true
            left end of clone RP11-419C13 is at 55152 in this sequence. The
            true right end of clone RP11-475B22 is at 55157 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .149849
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-141E7"
                     /clone_lib="RPCI-11.1"
     repeat_region   6. .280
                     /note="L1M1 repeat: matches -1378. .-1092 of consensus"
     repeat_region   302. .635
                     /note="L1 repeat: matches 2903. .3224 of consensus"
     repeat_region   637. .1056
                     /note="LTR12 repeat: matches 1. .437 of consensus"
     repeat_region   1057. .1168
                     /note="L1 repeat: matches 2794. .2906 of consensus"
     repeat_region   1178. .1656
                     /note="L1 repeat: matches 2365. .2486 of consensus"
     repeat_region   1882. .2658
                     /note="L1M3b repeat: matches 1331. .1712 of consensus"
     repeat_region   2659. .2960
                     /note="AluJo repeat: matches 1. .296 of consensus"
     repeat_region   2961. .3582
                     /note="L1M3b repeat: matches 699. .1331 of consensus"
     repeat_region   3818. .4053
                     /note="LTR12 repeat: matches 16. .251 of consensus"
     repeat_region   4054. .4168
                     /note="LTR25 repeat: matches 608. .721 of consensus"
     repeat_region   4169. .4518
                     /note="LTR12 repeat: matches 250. .671 of consensus"
     repeat_region   4759. .5202
                     /note="L1M3c repeat: matches 18. .466 of consensus"
     repeat_region   6627. .6915
                     /note="AluSc repeat: matches 3. .296 of consensus"
     repeat_region   10257. .10419
                     /note="L1P3 repeat: matches 5439. .5608 of consensus"
     repeat_region   10420. .10810
                     /note="L1PA7 repeat: matches 5749. .6141 of consensus"
     repeat_region   10857. .11277
                     /note="L1MEc repeat: matches 1619. .2059 of consensus"
     repeat_region   11332. .11499
                     /note="L1MB6 repeat: matches 5825. .6001 of consensus"
     repeat_region   11538. .11823
                     /note="AluY repeat: matches 7. .294 of consensus"
     repeat_region   11943. .12008
                     /note="L1 repeat: matches 3515. .3580 of consensus"
     repeat_region   12009. .12309
                     /note="AluSc repeat: matches 1. .306 of consensus"
     repeat_region   12310. .12941
                     /note="L1 repeat: matches 3580. .4208 of consensus"
     repeat_region   13055. .13820
                     /note="L1 repeat: matches 3658. .4418 of consensus"
     repeat_region   13776. .15974
                     /note="L1MC1 repeat: matches 3658. .5866 of consensus"
     repeat_region   16050. .16359
                     /note="AluSg repeat: matches 1. .308 of consensus"
     repeat_region   19074. .19153
                     /note="40 copies 2 mer at 75% conserved"
     repeat_region   19078. .19153
                     /note="19 copies 4 mer atat 76% conserved"
     repeat_region   19344. .19564
                     /note="MIR repeat: matches 48. .260 of consensus"
     repeat_region   21244. .21493
                     /note="MIR repeat: matches 2. .249 of consensus"
     repeat_region   21524. .21596
                     /note="MIR repeat: matches 91. .171 of consensus"
     repeat_region   22467. .23005
                     /note="LTR35 repeat: matches 73. .616 of consensus"
     repeat_region   23525. .23759
                     /note="L2 repeat: matches 1979. .2201 of consensus"
     repeat_region   23760. .23978
                     /note="MER58A repeat: matches 2. .221 of consensus"
     repeat_region   23979. .24040
                     /note="L2 repeat: matches 2201. .2264 of consensus"
     repeat_region   24238. .24466
                     /note="MIR repeat: matches 3. .256 of consensus"
     repeat_region   25207. .25340
                     /note="L2 repeat: matches 2607. .2749 of consensus"
     repeat_region   25725. .25756
                     /note="8 copies 4 mer tcct 90% conserved"
     repeat_region   25773. .26065
                     /note="AluSx repeat: matches 14. .307 of consensus"
     repeat_region   26535. .27027
                     /note="L1MC4 repeat: matches 7466. .7964 of consensus"
     repeat_region   27302. .27385
                     /note="L1MC/D repeat: matches 5787. .5870 of consensus"
     repeat_region   27379. .27977
                     /note="L1MC4 repeat: matches 6422. .7035 of consensus"
     repeat_region   28144. .28183
                     /note="20 copies 2 mer ct 77% conserved"
     repeat_region   28236. .28468
                     /note="AluJo repeat: matches 1. .288 of consensus"
     repeat_region   28799. .28830
                     /note="16 copies 2 mer ac 93% conserved"
     repeat_region   29111. .29698
                     /note="L1MC/D repeat: matches 4895. .5600 of consensus"
     repeat_region   29688. .30418
                     /note="L1MA5 repeat: matches 5176. .5877 of consensus"
     repeat_region   30422. .30670
                     /note="L1MA5 repeat: matches 6056. .6299 of consensus"
     repeat_region   30865. .33326
                     /note="L1PA4 repeat: matches 3686. .6146 of consensus"
     repeat_region   33588. .33695
                     /note="54 copies 2 mer ta 81% conserved"
     repeat_region   33592. .33695
                     /note="26 copies 4 mer tata 82% conserved"
     repeat_region   33702. .33793
                     /note="23 copies 4 mer atat 68% conserved"
     repeat_region   37168. .37352
                     /note="AluJb repeat: matches 91. .293 of consensus"
     repeat_region   39051. .39132
                     /note="MIR repeat: matches 184. .262 of consensus"
```

```
     repeat_region   39849. .40145
                     /note="AluSx repeat: matches 1. .301 of consensus"
     repeat_region   40659. .40870
                     /note="L2 repeat: matches 820. .1020 of consensus"
     repeat_region   40930. .41033
                     /note="L2 repeat: matches 1016. .1115 of consensus"
     repeat_region   41447. .41913
                     /note="L2 repeat: matches 1639. .2136 of consensus"
     repeat_region   41936. .42773
                     /note="L1PA13 repeat: matches 5355. .6148 of consensus"
     repeat_region   42774. .43068
                     /note="AluSq repeat: matches 1. .312 of consensus"
     repeat_region   43069. .43636
                     /note="L1PA13 repeat: matches 4773. .5355 of consensus"
     repeat_region   43637. .43890
                     /note="pTR5 repeat: matches 743. .998 of consensus"
     repeat_region   43984. .44783
                     /note="LTR12 repeat: matches 18. .671 of consensus"
     repeat_region   44784. .45860
                     /note="L1PA13 repeat: matches 3679. .4778 of consensus"
     repeat_region   45887. .46370
                     /note="L1P repeat: matches 3121. .3665 of consensus"
     repeat_region   46391. .46479
                     /note="L2 repeat: matches 2155. .2243 of consensus"
     repeat_region   46740. .47159
                     /note="30 copies 14 mer 55% conserved"
     repeat_region   46770. .47199
                     /note="215 copies 2 mer aa 54% conserved"
     repeat_region   49260. .49623
                     /note="L1MA1 repeat: matches 5900. .6274 of consensus"
     repeat_region   50874. .51178
                     /note="AluSq repeat: matches 1. .303 of consensus"
     repeat_region   51361. .51638
                     /note="AluJb repeat: matches 1. .286 of consensus"
     repeat_region   51852. .51887
                     /note="9 copies 4 mer acac 94% conserved"
     repeat_region   52037. .52248
                     /note="MIR repeat: matches 19. .251 of consensus"
     repeat_region   53981. .54137
                     /note="L1 repeat: matches 5037. .5196 of consensus"
     repeat_region   54634. .54661
                     /note="14 copies 2 mer ag 100% conserved"

  Query Match             0.2%;  Score 21;  DB 9;  Length 149849;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2383 AACAAGCAAATGGAAAACATT 2403
             |||||||||||||||||||||
Db     100660 AACAAGCAAATGGAAAACATT-100640


RESULT 94
AC091915
LOCUS       AC091915               150965 bp    DNA     linear   PRI 21-DEC-2001
DEFINITION  Homo sapiens chromosome 5 clone RP11-224A9, complete sequence.
ACCESSION   AC091915
VERSION     AC091915.4  GI:17975242
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 150965)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 150965)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 150965)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 150965)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Dec 21, 2001 this sequence version replaced gi:17149379.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.8% of Sequence;
            Estimated Total Number of Errors is 0.3.
FEATURES             Location/Qualifiers
     source          1. .150965
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-224A9"
BASE COUNT    45476 a  26753 c  28930 g  49806 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 9;  Length 150965;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2372 AAATAAATAAAAACAAGCAAA 2392
             |||||||||||||||||||||
Db     123426 AAATAAATAAAAACAAGCAAA 123446


RESULT 95
HS404K8/c
LOCUS       HS404K8                151236 bp    DNA     linear   PRI 04-MAR-2003
DEFINITION  Human DNA sequence from clone RP3-404K8 on chromosome 6p22.2-22.3,
            complete sequence.
ACCESSION   AL023883
VERSION     AL023883.6  GI:4775332
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151236)
  AUTHORS   Nickerson,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 11, 1999 this sequence version replaced gi:4741469.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
```

```
            Contact: humquery@sanger.ac.uk
            ---------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            RP3-404K8 is from the library RPCI-3 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .151236
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="p22.2-22.3"
                     /clone="RP3-404K8"
                     /clone_lib="RPCI-3"
BASE COUNT    49397 a  29800 c  28255 g  43784 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 9;  Length 151236;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4150 CTGTTTTAAAAATGGAGAAAG 4170
              |||||||||||||||||||||
Db      72778 CTGTTTTAAAAATGGAGAAAG 72758


RESULT 96
AC104322
LOCUS       AC104322              151428 bp    DNA     linear   PLN 14-SEP-2002
DEFINITION  Oryza sativa chromosome 10 BAC OSJNBb0005F01 genomic sequence,
            complete sequence.
ACCESSION   AC104322
VERSION     AC104322.4  GI:22795302
KEYWORDS    HTG.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 151428)
  AUTHORS   Buell,C.R., Yuan,Q., Ouyang,S., Liu,J., Gansberger,K., Jones,K.M.,
            Overton II,L.L., Tsitrin,T., Kim,M.M., Bera,J.J., Jin,S.S.,
            Fadrosh,D.W., Tallon,L.J., Koo,H., Zismann,V., Hsiao,J., Blunt,S.,
            Vanaken,S.S., Riedmuller,S.B., Utterback,T.T., Feldblyum,T.V.,
            Yang,Q.Q., Haas,B.J., Suh,B.B., Peterson,J.J., Quackenbush,J.,
            White,O., Salzberg,S.L. and Fraser,C.M.
  TITLE     Oryza sativa chromosome 10 BAC OSJNBb0005F01 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151428)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 151428)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   4  (bases 1 to 151428)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-SEP-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
COMMENT     On Sep 12, 2002 this sequence version replaced gi:19881809.
            Address all correspondence to:rice@tigr.org

            BAC clone OSJNBb0005F01 is from Oryza sativa chromosome 10
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Fgenesh (http://www.softberry.com/),
            genscan and Genscan+ (Chris Burge,
            http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
            http://genemark.biology.gatech.edu/GeneMark/), and GeneSplicer
            (Mihaela Pertea and Steven Salzberg, contact mpertea@tigr.org),
            searches of the complete sequence against a peptide database and
            the plant EST database at TIGR (http://www.tigr.org/tdb/tgi.shtml).
            Annotated genes are named to indicate the level of evidence for
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes without significant peptide
            similarity but with EST similarity are named as unknown proteins.
            Genes without protein or EST similarity, that are predicted by more
            than two gene prediction programs over most of their length are
            annotated as hypothetical proteins. Genes encoding tRNAs are
            predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
            identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html).
            This BAC overlaps with rice BAC OSJNBa0040E17 (AC097278) and
            OSJNBa0034E15 (AC105377).
FEATURES             Location/Qualifiers
     source          1. .151428
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="10"
                     /map="R10430S"
                     /clone="OSJNBb0005F01"
                     /note="japonica cultivar-group"
     gene            complement(323. .4813)
                     /gene="OSJNBb0005F01.1"
                     /note="similar to helicase 2 GB:AAK70786 GI:14591881
                     (Cydia pomonella granulovirus)"
     mRNA            complement(join(<323. .3549,3648. .4553,4628. .4681,
                     4756. .>4813))
                     /gene="OSJNBb0005F01.1"
     CDS             complement(join(<323. .3549,3648. .4553,4628. .4681,
                     4756. .4813))
                     /gene="OSJNBb0005F01.1"
                     /codon_start=1
                     /product="putative helicase, 3'-partial"
                     /protein_id="AAN09850.1"
                     /db_xref="GI:22857576"
                     /translation="METTHEDDINDECHMFAGEDMDFDDDDDDEYYMFAGGDGEDEED
                     IDLNEIEQDSVEHDPYDHVYSNIPRNTHVLKPTANCKHCGAKKFQYETNGFCCRGGKI
                     KLSNLETPPELMRLWSSMDSDATHFRDNIRFFNGHFSFTTLGVSLDERYTNMRSGVYT
                     FRAHGQIYHNIHSFGQRDNAPEHLELYFYDDDPSLNHRFRRSPSLDQEVVRRLVEVLR
                     GNPYSQTFRSLAQADDLEEYCVTLNLDNRLDQRRYNVPVTSEVAAVWVEGNELRTHFE
                     RSVVLYGNNNTKYSIQSYYGCYDPLSYPLFFPKGELGWHPEIPKVGVSIEDVIASRGN
                     NHADSDSNSRLCVSVRDYYCYKFQMRRGIFNPLLYGKRLFQQFAVDMYIKVESTRLDF
                     IRRHQVEIRADLYKGVVDSIHAGESRASQVGKRTVLPASFIGGNRDMKRRYMDAMALV
                     QKYGKPDIFLTMTCNPNWEEIIGNLEPGQTPQDRPDLIVRVFRAKLEDLKKQLLEKHI
                     LGKVIAYAYVVEFQKRGLPHAHFLLIMDGKYKLTSPEQYDCIISAELPNKHKYPELYE
                     MVVKHMMHGPCGTLNRKNVCMQDGSCKNRYPRAFNAVTINGKDSYPVYRRRDNKRCAK
                     VRKQMLDNRWVVPYNPYLLRMYNCHINVEVCSSIKAVKYLFKYIYKGHDRASVTLGEA
                     DSNGNIDEIQQYRDARWVTPPEALWRIYSFDLSKVHPPVKQLQLHLPNMHMVSFRAGQ
                     DLHDVVERNGVEKSMLTEYFEANKVNEDARGILYKDFPEAFTWQAGPKVWQRRKRRIT
                     QIGRIVTAHPAEGERYYLRVLLTHVTGATSFEHLRTVDGEVCSSFREAAEQRGLIEAD
                     NTLDECLTKAEVFQMPSSLRRLFATILVFCEPSDVLGLWNKHLEGMLDDYRRSHTCPR
                     TIEQMVLLDIRNMLQSMGKEITSFPLPEIDESYDTSGSEPREIFEESTIEVDHGHMDL
                     SSSLNPEQRCAYDEILSAIEGGQGGLFFVDGPGGTGKTFLYKALLATIRGQGKIAVAT
                     ATSGVAASIMLGGRTAHSRFKIPLNIDDGGVCNFTKQSGTAKLLQRASLIIWDEASMT
                     KRQAVEALDRSMRDIMDRPDVPFGGKTVVFGGDFRQVLPVVRKGTRPQITDATLRKSY
                     LWDCMRQLRLVTNMRAQNDRWFAEFLLRVGNGIEEACDDGYIRLPDEICVPCTGNDDD
                     LNNLIDNVFPMLDANLADPNYITSRAILSTRNEYVDQINMKMIDRFRGEEMLYHSFDR
                     AEDDPHNYYPPEFLNSLTPNGLPPHILKLKINCPVILLRNIDPANGLCNGTRLVVRGF
```

```
                QRNTIDAEIVLGQHAGKRVFLPRIPLCPSDDEMFPFRFKRKQFPVRLSFAMTINKAQG
                QTIPNVGVYLPDPVFSHGQLYVALSRATARMNIKILA"
repeat_region   5786. .5875
                /rpt_family="tandem"
gene            6350. .6556
                /gene="OSJNBb0005F01.2"
                /note="predicted by fgenesh"
mRNA            <6350. .>6556
                /gene="OSJNBb0005F01.2"
CDS             6350. .6556
                /gene="OSJNBb0005F01.2"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAN09851.1"
                /db_xref="GI:22857577"
                /translation="MAPWGGSPSRQRGGGGGFGRHGFFAGEGRRRRRRAGATARKKQR
                GRAGAAAAAGDGWRGSKGGAATLR"
repeat_region   6432. .6547
                /rpt_family="(CGG)n"
repeat_region   6462. .6525
                /rpt_family="tandem"
repeat_region   6586. .6671
                /rpt_family="(CGG)n"
repeat_region   6684. .6759
                /rpt_family="(GA)n"
repeat_region   7089. .7155
                /rpt_family="tandem"
repeat_region   complement(7323. .7366)
                /rpt_family="(GA)n"
gene            7513. .7974
                /gene="OSJNBb0005F01.3"
                /note="predicted by fgenesh"
mRNA            <7513. .>7974
                /gene="OSJNBb0005F01.3"
CDS             7513. .7974
                /gene="OSJNBb0005F01.3"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAN09853.1"
                /db_xref="GI:22857579"
                /translation="MSLARTAEPPWQPQRRCGGALALPLRGHDEVEEAASCASYRGQR
                GRGKAMEVRDEMHNAAGLREDGRGSGDILLPSLHCLPPHVAPLLSPFSHIGAWHHSGS
                EAAWSASATGHRGMLRLRCQPPRRALPQPPPSPEQAPSPQRHCGCHGDAAP"
repeat_region   8250. .8285
                /rpt_family="(GA)n"
repeat_region   complement(8329. .8360)
                /rpt_family="(TAGA)n"
gene            complement(9582. .12232)
                /gene="OSJNBb0005F01.5"
                /note="predicted by fgenesh"
mRNA            complement(join(<9582. .9695,9775. .10057,10138. .10308,
                10386. .11230,11716. .11751,11828. .>12232))
                /gene="OSJNBb0005F01.5"
CDS             complement(join(9582. .9695,9775. .10057,10138. .10308,
                10386. .11230,11716. .11751,11828. .12232))
                /gene="OSJNBb0005F01.5"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAN09854.1"
                /db_xref="GI:22857580"
                /translation="MGNCWKNFKTTLVTVYVLNPAQPEPFVKYPFITKPVWDEFHAAK
                SMNESREKSQTYRDLQARNLDPHRLGTGGYAGKQAEWDKEDEAAAESNTQPVLGDIPV
                QRARNWARARVKKNRDGTLSFPNPEDQAVYQKIVELNAERQASQEEPTSCTLVIRVTP
                GFAIPAAEGQAFKPTPKTRVHGAQLQAGNAKVQVDLVKPDWVGYTIPHPPNDEILTLG
                AARGTFIQWPKHSIEINITPRPAPSSRPPPTRPHQTVVSLPPAVEQRDEDLQLQYDTD
                FGDDGMEVDSRPHLPPPAKRSKRAKSSPPKLDTRRKAAGMGRGKIKVPLVPKKLDLGK
                APVAPPKPPAEFTLGMPLVGDDALFKMGPACKELHGYYMEKSNARRKNRETSMLGQHD
                GQPFLGPTAFIAVDFKDLWDLYRVRAIDTNLLKCYSLLTWKHVHRKAPHVALLDPAVD
                NETTLKNDRANMVAYIKDCLFARQDKDFIMCAYNQQRHWILLVITPKWSLVHYLNSNI
                KPEIYDWSAIESALNEAWDQYVARGGRHKDGHPKLGHRKDFPIRQQVGDQCGFHVCHN
                MRSFAEKVTLLDPEEFISRIPEINFEIIREELAQFILDDILSPKGMFRVKS"
repeat_region   11495. .11519
                /rpt_family="tandem"
repeat_region   13653. .14006
                /rpt_family="tandem"
repeat_region   14237. .14517
                /rpt_family="tandem"
repeat_region   14246. .14504
                /rpt_family="tandem"
repeat_region   14747. .15018
                /rpt_family="tandem"
repeat_region   14747. .15006
                /rpt_family="tandem"
repeat_region   complement(15506. .15547)
                /rpt_family="AT_rich"
repeat_region   15614. .15747
                /rpt_family="tandem"
repeat_region   complement(15679. .15722)
                /rpt_family="AT_rich"
gene            complement(16106. .19100)
                /gene="OSJNBb0005F01.6"

  Query Match             0.2%;  Score 21;  DB 8;  Length 151428;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4714 ATTAAAAAAAATCCACATGTA 4734
              |||||||||||||||||||||
Db      59762 ATTAAAAAAAATCCACATGTA 59782


RESULT 97
AC112229
LOCUS       AC112229              154279 bp    DNA     linear   PRI 24-OCT-2002
DEFINITION  Homo sapiens BAC clone RP13-1039J1 from 2, complete sequence.
ACCESSION   AC112229
VERSION     AC112229.4  GI:22758607
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 154279)
  AUTHORS   The Sanger Centre and The Washington University Genome Sequencing
            Center.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 154279)
  AUTHORS   Abbott,S., Strowmatt,C., Kozlowicz,A. and Dignan,G.
  TITLE     The sequence of Homo sapiens BAC clone RP13-1039J1
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 154279)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 154279)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 154279)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 154279)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
```

```
COMMENT     On Sep 7, 2002 this sequence version replaced gi:21747810.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_FH1039J01
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-13 Human Female BAC Library was constructed using improved
            cloning techniques developed by Kazutoyo Osoegawa. The library was
            generated by Baohui Zhao in our laboratory.  Construction was
            funded by a grant from the National Human Genome Research Institute
            (NHGRI, NIH) (#1R01RG01165-03). The library was generated according
            to the new NHGRI/DOE 'Guidance on Human Subjects in Large-Scale DNA
            Sequencing'.
            Female blood was obtained via a double-blind selection protocol.
            Female blood DNA was isolated from one randomly chosen donor (out
            of 10 female donors) and partially digested with a combination of
            EcoRI and EcoRI Methylase for library segments 1&2 or either MboI
            or DpnII for library segments 3&4. Size selected DNA was cloned
            into the pBACe3.6 vector between the EcoRI sites for library
            segments 1&2 or the BamHI sites for library segments 3&4. The
            ligation products were transformed into DH10B electrocompetent
            cells (BRL Life Technologies). The library has been arrayed into
            384-well microtiter dishes and also gridded onto 22x22cm nylon high
            density filters for screening by probe hybridization.

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES                Location/Qualifiers
     source             1. .154279
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="2"
                        /map="2"
                        /clone="RP13-1039J1"
                        /clone_lib="RPCI-13"
     repeat_region      1. .249
                        /rpt_family="ERV1"
     repeat_region      311. .415
                        /rpt_family="MIR"
     repeat_region      426. .639
                        /rpt_family="MER1_type"
     repeat_region      640. .1000
                        /rpt_family="MER2_type"
     repeat_region      1020. .1717
                        /rpt_family="L1"
     repeat_region      1718. .1897
                        /rpt_family="MER1_type"
     repeat_region      1898. .2023
                        /rpt_family="Alu"
     repeat_region      2024. .2068
                        /rpt_family="MER1_type"
     repeat_region      2069. .2357
                        /rpt_family="Alu"
     repeat_region      2358. .2483
                        /rpt_family="MER1_type"
     repeat_region      2477. .3141
                        /rpt_family="MER1_type"
     repeat_region      3142. .3273
                        /rpt_family="Alu"
     repeat_region      3274. .3359
                        /rpt_family="MER1_type"
     repeat_region      3828. .4106
                        /rpt_family="Alu"
     repeat_region      4842. .5124
                        /rpt_family="Alu"
     repeat_region      6067. .6317
                        /rpt_family="L1"
     repeat_region      6324. .6616
                        /rpt_family="Alu"
     repeat_region      6974. .7275
                        /rpt_family="Alu"
     repeat_region      7367. .7675
                        /rpt_family="Alu"
     repeat_region      8068. .8438
                        /rpt_family="MaLR"
     repeat_region      9775. .10722
                        /rpt_family="L1"
     repeat_region      11979. .12167
                        /rpt_family="MER2_type"
     repeat_region      12168. .12468
                        /rpt_family="Alu"
     repeat_region      12469. .12588
                        /rpt_family="MER2_type"
     repeat_region      12606. .12898
                        /rpt_family="Alu"
     repeat_region      12920. .13065
                        /rpt_family="Alu"
     repeat_region      13066. .13362
                        /rpt_family="Alu"
     repeat_region      13363. .13525
                        /rpt_family="Alu"
     repeat_region      14200. .14477
                        /rpt_family="Alu"
     repeat_region      15154. .15464
                        /rpt_family="Alu"
     repeat_region      15741. .15833
                        /rpt_family="L2"
     repeat_region      15834. .16130
                        /rpt_family="Alu"
     repeat_region      16626. .16928
                        /rpt_family="Alu"
     repeat_region      16947. .17418
                        /rpt_family="MER1_type"
     repeat_region      17459. .17768
                        /rpt_family="Alu"
     repeat_region      17818. .18125
                        /rpt_family="Alu"
     repeat_region      18126. .18177
                        /rpt_family="(TA)n"
     repeat_region      18218. .18314
                        /rpt_family="L2"
     repeat_region      18472. .18528
                        /rpt_family="L2"
     repeat_region      18847. .18946
                        /rpt_family="L1"
     repeat_region      18951. .19074
                        /rpt_family="Alu"
     repeat_region      19551. .20490
                        /rpt_family="L1"
     repeat_region      20585. .20891
                        /rpt_family="Alu"
```

```
     repeat_region   20941. .21066
                     /rpt_family="Alu"
     repeat_region   21116. .21388
                     /rpt_family="Tip100"
     repeat_region   21540. .21640
                     /rpt_family="ERVL"
     repeat_region   21735. .21891
                     /rpt_family="MER2_type"
     repeat_region   22101. .22252
                     /rpt_family="MER1_type"
     repeat_region   22683. .23060

  Query Match             0.2%;  Score 21;  DB 9;  Length 154279;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4011 AAGGATCCTGGCCAAGGATCA 4031
              |||||||||||||||||||||
Db     118790 AAGGATCCTGGCCAAGGATCA 118810


RESULT 98
AC097358
LOCUS       AC097358               154606 bp    DNA     linear   PRI 24-MAY-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-157F20, complete sequence.
ACCESSION   AC097358
VERSION     AC097358.2  GI:21166205
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 154606)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Raymond,C. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 154606)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 154606)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Raymond,C. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAY-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On May 24, 2002 this sequence version replaced gi:16152308.
            -------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
            -------------- Project Information
                Center project name: chr-3
                Center clone name: RP11-157F20 (bc0274)
            -------------- Summary Statistics
                Sequencing vector: plasmid; L08752; 100% of reads
                Chemistry: Dye-terminator ET; 96% of reads
                Chemistry: Dye-terminator Big Dye; 4% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 154096 bases at least Q40
                Consensus quality: 154548 bases at least Q30
                Consensus quality: 154600 bases at least Q20
                Insert size: 154606; sum-of-contigs
                Quality coverage: 8.5x in Q20 bases; sum-of-contigs
            ----------
            Overlapping Sequences:
                5': RP11-263I16 (UWGC:bc0327) AC092050, 123826-bp overlap
                3': RP11-241K3 (UWGC:bc0319) AC093928
            --------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.

            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            --------------------
            Sequence Validation:
             This sequence has been validated by Multiple Complete Digest
             fingerprinting. Comparison of the experimentally derived digest
             fragments with sequence-predicted fragments is given below.
             The electronically-digested sequence consists of both insert and
             vector, in order to accurately represent the entire circular BAC.
             Small fragments below a variable cutoff (approximately 400-800 bp)
             are not resolved in the fingerprint and hence do not appear
             in the table. There are no significant remaining discrepancies
             between the experimental and predicted values. Uniquely ordered
             fragments are separated by dashed lines.
```

HindIII		EcoRI		BglII	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
3231	3121	8696	8585	3265	3145
6382	6538	6	<800	2067	2073
512	<800	7382	7337	6892	7026
449	<800	4493	4449	16	<800
14536	14523	4776	4809	679	<800
2492	2542	10400	10102	5157	5118
710	<800	635	<800	695	<800
1773	1776	6842	6855	4557	4598
1817	1776	30273	30419	172	<800
10199	10201	337	<800	3085	3145
3565	3541	3167	3155	9859	9864
2681	2723	2854	2931	3075	3145
708	<800	117	<800	310	<800
3604	3541	4865	4809	1004	1023
664	<800	2210	2213	35	<800
8301	8408	2331	2370	4244	4179
1180	1177	1318	1293	413	<800
1706	1668	2938	2931	4845	4857
3801	3792	584	<800	614	<800

3437	3541	4877	5154	5589	5568
1650	1668	831	840	575	<800
5467	5467	6053	6013	6678	6590
5017	4986	5868	5819	2413	2452
10789	10755	8515	8585	810	779
947	965	5548	5819	8618	8697
1734	1668	2823	2799	6454	6590
1183	1177	2018	1965	6595	6590
4484	4282	318	<800	2965	2992
559	<800	1447	1422	5813	5792
5912	5911	4264	4257	757	779
4341	4282	8	<800	608	<800
310	<800	1098	1079	3318	3325
464	<800	1916	1965	4121	4179
636	<800	2021	1965	4506	4437
70	<800	5229	5450	1691	1729
18	<800	13004	12667	3662	3704
1180	1177	580	<800	2284	2325
427	<800	2663	2648	1519	1510
3414	3541			3887	3864
6212	6170			4692	4598
4183	4282			7645	7623
2010	2033			9033	8940
826	845			11185	11154
3080	3121			1184	1192
4210	4282			3897	3864
2047	2033			1822	1729
9629	9493				
10758	10755				

```
FEATURES             Location/Qualifiers
     source          1. .154606
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-157F20"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    76623. .76639
                     /note="Single subclone region"
     misc_feature    76783. .77114
                     /note="Single subclone region"
BASE COUNT    44308 a  35892 c  34096 g  40310 t
ORIGIN
```

```
  Query Match             0.2%;  Score 21;  DB 9;  Length 154606;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5316 GGGCAGCATATTTATTTTTAT 5336
              |||||||||||||||||||||
Db     110742 GGGCAGCATATTTATTTTTAT 110762


RESULT 99
AC074087/c
LOCUS       AC074087               156083 bp    DNA     linear   PRI 07-NOV-2001
DEFINITION  Homo sapiens BAC clone RP11-26C10 from 4, complete sequence.
ACCESSION   AC074087
VERSION     AC074087.4  GI:15638721
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156083)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 156083)
  AUTHORS   Kang,K., Cordum,H., Doebber,A. and Dignan,G.
  TITLE     The sequence of Homo sapiens BAC clone RP11-26C10
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 156083)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 156083)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 156083)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Sep 18, 2001 this sequence version replaced gi:14029076.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0026C10
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
```

McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is AC021417; the clone sequenced to the right is RP11-691E1. Actual start of this clone is at base position 1 of RP11-26C10; actual end is at base position 156083 of RP11-26C10.

There are polymorphic base pair differences in the overlap between the clone RP11-26C10 and RP11-691E1.

Data from AC019311 (H_NH0691E01) was used to finish this clone, AC074087 (H_NH0026C10).

The location of this clone is unknown.

FEATURES Location/Qualifiers
source 1. .156083
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="unknown"
/clone="RP11-26C10"
/clone_lib="RPCI-11"
repeat_region 66. .846
/rpt_family="L1"
repeat_region 494. .514
/rpt_family="AT_rich"
repeat_region 1194. .1764
/rpt_family="MaLR"
repeat_region 1767. .2080
/rpt_family="L1"
repeat_region 2273. .2309
/rpt_family="(TTTA)n"
repeat_region 2392. .3472
/rpt_family="L2"
repeat_region 4722. .4750
/rpt_family="AT_rich"
repeat_region 4941. .5602
/rpt_family="ERV1"
repeat_region 7562. .7789
/rpt_family="L1"
repeat_region 7929. .8229
/rpt_family="Alu"
repeat_region 8643. .8676
/rpt_family="(A)n"
repeat_region 8797. .9091
/rpt_family="Alu"
repeat_region 9842. .10143
/rpt_family="Alu"
repeat_region 10298. .10489
/rpt_family="L1"
repeat_region 10547. .10592
/rpt_family="(TA)n"
repeat_region 10930. .11284
/rpt_family="ERV1"
repeat_region 11496. .11601
/rpt_family="(TCTCC)n"
repeat_region 11613. .11878
/rpt_family="Alu"
repeat_region 11952. .11972
/rpt_family="AT_rich"
repeat_region 12965. .13037
/rpt_family="MIR"
repeat_region 13038. .13370
/rpt_family="MaLR"
repeat_region 13371. .13438
/rpt_family="MIR"
repeat_region 13489. .13523
/rpt_family="AT_rich"
repeat_region 13594. .14448
/rpt_family="MER2_type"
repeat_region 14445. .14605
/rpt_family="MER2_type"
repeat_region 14629. .14756
/rpt_family="L1"
repeat_region 14894. .14971
/rpt_family="MIR"
repeat_region 15259. .15705
/rpt_family="ERVL"
repeat_region 16656. .17103
/rpt_family="L2"
repeat_region 17435. .18748
/rpt_family="L1"
repeat_region 19128. .19418
/rpt_family="Alu"
repeat_region 19424. .20050
/rpt_family="L2"
repeat_region 20287. .20339
/rpt_family="AT_rich"
repeat_region 21411. .21456
/rpt_family="GA-rich"
repeat_region 22595. .23137
/rpt_family="ERVL"
repeat_region 23342. .23513
/rpt_family="L1"
repeat_region 23514. .25349
/rpt_family="L1"
repeat_region 26619. .26698
/rpt_family="L2"
repeat_region 26699. .26996
/rpt_family="Alu"
repeat_region 26997. .27058
/rpt_family="L2"
repeat_region 27066. .27240
/rpt_family="MIR"
repeat_region 27321. .27611
/rpt_family="MaLR"
repeat_region 27652. .27732
/rpt_family="MIR"
repeat_region 28959. .28982
/rpt_family="(T)n"
repeat_region 29516. .29603
/rpt_family="L2"
repeat_region 29921. .30041
/rpt_family="MER1_type"
repeat_region 30493. .30760
/rpt_family="MER1_type"
repeat_region 30856. .31015
/rpt_family="MER1_type"
repeat_region 31044. .31131
/rpt_family="CT-rich"
repeat_region 32437. .32512
/rpt_family="AcHobo"
repeat_region 32964. .33045
/rpt_family="MER1_type"
repeat_region 35014. .35611
/rpt_family="L2"

Query Match 0.2%; Score 21; DB 9; Length 156083;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7488 ACCTAAATAAGCTCACAGAAG 7508
||||||||||||||||||||

```
Db      49320 ACCTAAATAAGCTCACAGAAG 49300

RESULT 100
AC027163
LOCUS       AC027163              156690 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 6 clone RP11-607F22 map 6, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC027163
VERSION     AC027163.2  GI:10567962
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156690)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 6, clone RP11-607F22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156690)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 156690)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 4, 2000 this sequence version replaced gi:7331476.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
               Center project name: L6511
               Center clone name: 607_F_22
            -------------- Summary Statistics
               Sequencing vector: M13; M77815; 100% of reads
               Chemistry: Dye-terminator Big Dye; 100% of reads
               Assembly program: Phrap; version 0.960731
               Consensus quality: 150593 bases at least Q40
               Consensus quality: 153554 bases at least Q30
               Consensus quality: 154789 bases at least Q20
               Insert size: 153000; agarose-fp
               Insert size: 155490; sum-of-contigs
               Quality coverage: 5.1 in Q20 bases; agarose-fp
               Quality coverage: 5.0 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1720: contig of 1720 bp in length
            *     1721     1820: gap of 100 bp
            *     1821     4662: contig of 2842 bp in length
            *     4663     4762: gap of 100 bp
            *     4763     8750: contig of 3988 bp in length
            *     8751     8850: gap of 100 bp
            *     8851    15813: contig of 6963 bp in length
            *    15814    15913: gap of 100 bp
            *    15914    23306: contig of 7393 bp in length
            *    23307    23406: gap of 100 bp
            *    23407    33170: contig of 9764 bp in length
            *    33171    33270: gap of 100 bp
            *    33271    44422: contig of 11152 bp in length
            *    44423    44522: gap of 100 bp
            *    44523    81260: contig of 36738 bp in length
            *    81261    81360: gap of 100 bp
            *    81361    95970: contig of 14610 bp in length
            *    95971    96070: gap of 100 bp
            *    96071   106973: contig of 10903 bp in length
            *   106974   107073: gap of 100 bp
            *   107074   122796: contig of 15723 bp in length
            *   122797   122896: gap of 100 bp
            *   122897   144366: contig of 21470 bp in length
            *   144367   144466: gap of 100 bp
            *   144467   156690: contig of 12224 bp in length.
FEATURES             Location/Qualifiers
     source          1. .156690
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6"
                     /clone="RP11-607F22"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1720
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    1821. .4662
                     /note="assembly_fragment"
     misc_feature    4763. .8750
                     /note="assembly_fragment"
     misc_feature    8851. .15813
                     /note="assembly_fragment"
     misc_feature    15914. .23306
                     /note="assembly_fragment"
```

```
     misc_feature    23407. .33170
                     /note="assembly_fragment"
     misc_feature    33271. .44422
                     /note="assembly_fragment"
     misc_feature    44523. .81260
                     /note="assembly_fragment"
     misc_feature    81361. .95970
                     /note="assembly_fragment"
     misc_feature    96071. .106973
                     /note="assembly_fragment"
     misc_feature    107074. .122796
                     /note="assembly_fragment"
     misc_feature    122897. .144366
                     /note="assembly_fragment"
     misc_feature    144467. .156690
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    47316 a  29721 c  29869 g  48582 t   1202 others
ORIGIN

  Query Match             0.2%;  Score 21;  DB 2;  Length 156690;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4150 CTGTTTTAAAAATGGAGAAAG 4170
              |||||||||||||||||||||
Db      71465 CTGTTTTAAAAATGGAGAAAG 71485


Search completed: August 22, 2003, 10:12:45
Job time : 31591 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 21, 2003, 21:03:57 ; Search time 75.962 Seconds
(without alignments)
5810.577 Million cell updates/sec

Title: US-09-831-000-1_COPY_1_1000
Perfect score: 1000
Sequence: 1 gatcgggaaaacgcgagggg.....gtgcacggccaagttacaag 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	127.6	12.8	35100	2	US-08-770-379-17	Sequence 17, Appl
c	2	127.6	12.8	35100	3	US-08-757-669A-17	Sequence 17, Appl
c	3	127.6	12.8	35100	4	US-09-230-371A-17	Sequence 17, Appl
c	4	101.8	10.2	11558	5	PCT-US93-06251-23	Sequence 23, Appl
	5	77.4	7.7	1830121	4	US-09-557-884-1	Sequence 1, Appli
	6	77.4	7.7	1830121	4	US-09-643-990A-1	Sequence 1, Appli
c	7	61.8	6.2	1771	2	US-08-533-669A-7	Sequence 7, Appli
c	8	61.8	6.2	1771	2	US-08-511-872-1	Sequence 1, Appli
c	9	61.8	6.2	1771	4	US-09-183-861-7	Sequence 7, Appli
c	10	61.8	6.2	1771	4	US-09-022-765-7	Sequence 7, Appli
c	11	61.8	6.2	1771	4	US-09-551-974A-7	Sequence 7, Appli
c	12	60.8	6.1	45546	4	US-09-146-053-6	Sequence 6, Appli
c	13	58	5.8	3252	4	US-08-476-102A-4	Sequence 4, Appli
c	14	58	5.8	3252	5	PCT-US96-05320A-714	Sequence 714, App
c	15	56.8	5.7	350	2	US-08-332-766A-20	Sequence 20, Appl
c	16	54.4	5.4	538	2	US-08-332-766A-24	Sequence 24, Appl
c	17	53.8	5.4	72604	4	US-09-268-992-7	Sequence 7, Appli
c	18	53.8	5.4	72604	4	US-09-657-474-7	Sequence 7, Appli
c	19	53.4	5.3	319608	4	US-09-539-333D-1	Sequence 1, Appli
c	20	53.4	5.3	319608	4	US-09-679-409-1	Sequence 1, Appli
c	21	53	5.3	1960	2	US-08-595-937A-1	Sequence 1, Appli
	22	52.6	5.3	7218	1	US-08-232-463-14	Sequence 14, Appl
c	23	51.8	5.2	289	3	US-09-007-005-17	Sequence 17, Appl
c	24	51.8	5.2	289	3	US-09-244-796-17	Sequence 17, Appl
c	25	50.8	5.1	1559	3	US-09-019-095A-7	Sequence 7, Appli
c	26	50.4	5.0	1682	3	US-09-318-443-7	Sequence 7, Appli
c	27	49.8	5.0	348	2	US-08-623-906A-14	Sequence 14, Appl
	28	49.6	5.0	90541	4	US-09-759-359A-3	Sequence 3, Appli
	29	49.2	4.9	72604	4	US-09-268-992-7	Sequence 7, Appli
	30	49.2	4.9	72604	4	US-09-657-474-7	Sequence 7, Appli
c	31	48.8	4.9	650	1	US-08-229-515A-13	Sequence 13, Appl
c	32	48.8	4.9	650	1	US-08-645-865-13	Sequence 13, Appl
	33	48.4	4.8	405	2	US-08-903-800A-3	Sequence 3, Appli
c	34	48.2	4.8	951	4	US-09-252-991A-16386	Sequence 16386, A
	35	48.2	4.8	1092	4	US-09-252-991A-16212	Sequence 16212, A
c	36	48.2	4.8	1311	4	US-09-252-991A-16493	Sequence 16493, A
	37	48.2	4.8	1530	4	US-09-252-991A-16005	Sequence 16005, A
	38	48	4.8	756	4	US-09-252-991A-6106	Sequence 6106, Ap
	39	48	4.8	912	4	US-09-252-991A-6186	Sequence 6186, Ap
c	40	48	4.8	915	4	US-09-252-991A-6445	Sequence 6445, Ap
c	41	48	4.8	1200	4	US-09-252-991A-6282	Sequence 6282, Ap
	42	48	4.8	1806	4	US-09-252-991A-6023	Sequence 6023, Ap
	43	47.8	4.8	3228	4	US-09-579-664B-5	Sequence 5, Appli
c	44	47	4.7	18443	3	US-09-078-294-6	Sequence 6, Appli
c	45	46.8	4.7	446	2	US-08-332-766A-26	Sequence 26, Appl
c	46	46.4	4.6	124	2	US-08-332-766A-37	Sequence 37, Appl
c	47	46.2	4.6	152331	3	US-09-128-155-16	Sequence 16, Appl
	48	46.2	4.6	176373	3	US-09-128-155-17	Sequence 17, Appl
c	49	45.2	4.5	248	3	US-09-007-005-32	Sequence 32, Appl
c	50	45.2	4.5	248	3	US-09-244-796-32	Sequence 32, Appl
c	51	45.2	4.5	277	3	US-09-007-005-3	Sequence 3, Appli
c	52	45.2	4.5	277	3	US-09-244-796-3	Sequence 3, Appli
c	53	45.2	4.5	1978	1	US-07-753-520B-2	Sequence 2, Appli
c	54	45.2	4.5	9115	1	US-07-753-520B-3	Sequence 3, Appli
c	55	45	4.5	275	4	US-08-585-593A-42	Sequence 42, Appl
c	56	43.8	4.4	1667	1	US-08-485-284A-1	Sequence 1, Appli
c	57	43.8	4.4	20966	4	US-09-776-976-7	Sequence 7, Appli
c	58	43.8	4.4	20966	4	US-09-909-547-7	Sequence 7, Appli
c	59	43.8	4.4	20966	4	US-09-569-852B-1	Sequence 1, Appli
	60	42.8	4.3	4223	3	US-08-845-258-7	Sequence 7, Appli
c	61	42.8	4.3	4223	3	US-08-845-258-45	Sequence 45, Appl
	62	42.8	4.3	4223	3	US-08-990-571-7	Sequence 7, Appli
c	63	42.8	4.3	4223	3	US-08-990-571-45	Sequence 45, Appl
	64	42.8	4.3	4223	4	US-08-723-142A-7	Sequence 7, Appli
c	65	42.8	4.3	4223	4	US-08-723-142A-45	Sequence 45, Appl
	66	42.8	4.3	4223	4	US-09-528-784A-7	Sequence 7, Appli
c	67	42.8	4.3	4223	4	US-09-528-784A-45	Sequence 45, Appl
	68	42.8	4.3	4223	4	US-09-569-098A-7	Sequence 7, Appli
c	69	42.8	4.3	4223	4	US-09-569-098A-45	Sequence 45, Appl
	70	42	4.2	460	2	US-08-903-800A-4	Sequence 4, Appli
c	71	42	4.2	465	4	US-09-252-991A-16275	Sequence 16275, A
c	72	42	4.2	2445	3	US-09-298-367B-9	Sequence 9, Appli
c	73	42	4.2	8501	3	US-09-298-367B-6	Sequence 6, Appli
	74	42	4.2	162450	4	US-09-345-882-1	Sequence 1, Appli
	75	41.8	4.2	152331	3	US-09-128-155-16	Sequence 16, Appl
c	76	41.6	4.2	202001	4	US-09-734-674-3	Sequence 3, Appli
	77	41.4	4.1	64467	4	US-09-803-671B-3	Sequence 3, Appli
	78	41	4.1	1227	2	US-08-903-800A-1	Sequence 1, Appli
c	79	40.8	4.1	291	1	US-07-922-723A-7	Sequence 7, Appli
c	80	40.8	4.1	291	1	US-07-799-828C-7	Sequence 7, Appli
c	81	40.8	4.1	291	1	US-08-074-275-7	Sequence 7, Appli
c	82	40.8	4.1	291	1	US-08-480-366-7	Sequence 7, Appli
c	83	40.8	4.1	291	2	US-07-952-277A-7	Sequence 7, Appli
	84	40.4	4.0	453	2	US-08-903-800A-2	Sequence 2, Appli
	85	40.4	4.0	1679	2	US-08-365-486A-11	Sequence 11, Appl
	86	40.4	4.0	1679	3	US-08-880-342-11	Sequence 11, Appl
c	87	40	4.0	44453	4	US-09-146-053-5	Sequence 5, Appli
c	88	39.2	3.9	15144	3	US-08-458-434A-6	Sequence 6, Appli
c	89	39	3.9	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	90	39	3.9	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	91	38.8	3.9	217	2	US-08-332-766A-4	Sequence 4, Appli
c	92	38.8	3.9	454	2	US-08-623-906A-6	Sequence 6, Appli
c	93	38.8	3.9	1431	4	US-09-690-454-21	Sequence 21, Appl
c	94	38.6	3.9	16389	4	US-09-741-154-3	Sequence 3, Appli
	95	38.6	3.9	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	96	38.6	3.9	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	97	38.2	3.8	243	1	US-07-922-723A-9	Sequence 9, Appli
c	98	38.2	3.8	243	1	US-07-799-828C-9	Sequence 9, Appli
c	99	38.2	3.8	243	1	US-08-074-275-9	Sequence 9, Appli
c	100	38.2	3.8	243	1	US-08-480-366-9	Sequence 9, Appli

c 101	38.2	3.8	243	2	US-07-952-277A-9	Sequence 9, Appli
102	38	3.8	35100	1	US-08-306-691B-19	Sequence 19, Appl
103	38	3.8	35100	5	PCT-US93-06251-19	Sequence 19, Appl
c 104	37.8	3.8	1798	4	US-08-687-691B-1	Sequence 1, Appli
105	37.8	3.8	3331	3	US-09-042-785A-1	Sequence 1, Appli
c 106	37.6	3.8	349	1	US-08-253-155A-14	Sequence 14, Appl
107	37.4	3.7	2042	3	US-09-063-237-2	Sequence 2, Appli
108	37.4	3.7	2043	1	US-07-914-281-1	Sequence 1, Appli
109	37.4	3.7	2043	1	US-08-393-246-1	Sequence 1, Appli
110	37.4	3.7	2043	1	US-08-273-411-4	Sequence 4, Appli
111	37.4	3.7	2043	1	US-08-525-058A-1	Sequence 1, Appli
112	37.4	3.7	2043	2	US-08-696-731-1	Sequence 1, Appli
113	37.4	3.7	2043	3	US-09-042-531-1	Sequence 1, Appli
114	37.4	3.7	2043	5	PCT-US91-00899-1	Sequence 1, Appli
c 115	37.4	3.7	12141	3	US-09-488-671-10	Sequence 10, Appl
116	37.2	3.7	445	2	US-08-332-766A-16	Sequence 16, Appl
c 117	37.2	3.7	450	4	US-09-252-991A-6365	Sequence 6365, Ap
c 118	37.2	3.7	520	1	US-08-268-348A-7	Sequence 7, Appli
c 119	37.2	3.7	520	1	US-08-268-348A-9	Sequence 9, Appli
120	37.2	3.7	1433	2	US-08-666-392A-1	Sequence 1, Appli
121	37.2	3.7	1433	3	US-09-199-926-1	Sequence 1, Appli
c 122	37.2	3.7	53332	4	US-09-801-861-3	Sequence 3, Appli
c 123	37	3.7	9168	4	US-09-687-731-11	Sequence 11, Appl
124	36.8	3.7	1641	4	US-09-252-991A-2066	Sequence 2066, Ap
125	36.8	3.7	1771	2	US-08-533-669A-7	Sequence 7, Appli
126	36.8	3.7	1771	2	US-08-511-872-1	Sequence 1, Appli
127	36.8	3.7	1771	4	US-09-183-861-7	Sequence 7, Appli
128	36.8	3.7	1771	4	US-09-022-765-7	Sequence 7, Appli
129	36.8	3.7	1771	4	US-09-551-974A-7	Sequence 7, Appli
c 130	36.8	3.7	2111	4	US-09-052-753B-6	Sequence 6, Appli
c 131	36.8	3.7	2112	4	US-09-252-991A-1820	Sequence 1820, Ap
132	36.8	3.7	2283	4	US-09-252-991A-2140	Sequence 2140, Ap
c 133	36.8	3.7	29629	4	US-09-729-995-3	Sequence 3, Appli
c 134	36.6	3.7	3919	1	US-08-301-722A-1	Sequence 1, Appli
135	36.4	3.6	202001	4	US-09-734-674-3	Sequence 3, Appli
c 136	36.2	3.6	460	4	US-09-635-872A-12	Sequence 12, Appl
c 137	36.2	3.6	460	4	US-09-636-077A-12	Sequence 12, Appl
138	36.2	3.6	6803	3	US-08-665-259-19	Sequence 19, Appl
139	36.2	3.6	6803	3	US-08-762-500-19	Sequence 19, Appl
c 140	36.2	3.6	7218	1	US-08-232-463-14	Sequence 14, Appl
c 141	35.8	3.6	301	2	US-08-332-766A-23	Sequence 23, Appl
142	35.8	3.6	1743	3	US-08-665-259-20	Sequence 20, Appl
143	35.8	3.6	1743	3	US-08-762-500-20	Sequence 20, Appl
144	35.8	3.6	1974	3	US-08-762-500-78	Sequence 78, Appl
145	35.6	3.6	801	4	US-09-252-991A-4974	Sequence 4974, Ap
c 146	35.6	3.6	1500	5	PCT-US93-08386-3	Sequence 3, Appli
c 147	35.6	3.6	1758	4	US-09-252-991A-4956	Sequence 4956, Ap
148	35.6	3.6	3607	1	US-08-647-351B-1	Sequence 1, Appli
149	35.6	3.6	3607	4	US-09-326-074-1	Sequence 1, Appli
c 150	35.4	3.5	287	2	US-08-332-766A-27	Sequence 27, Appl
151	35.4	3.5	394	2	US-08-623-906A-7	Sequence 7, Appli
152	35.2	3.5	434	2	US-08-332-766A-10	Sequence 10, Appl
c 153	35.2	3.5	735	4	US-09-252-991A-1894	Sequence 1894, Ap
c 154	35.2	3.5	808	3	US-08-998-416-461	Sequence 461, App
155	35.2	3.5	1682	3	US-09-318-443-7	Sequence 7, Appli
c 156	35	3.5	462	4	US-09-252-991A-10	Sequence 10, Appl
c 157	35	3.5	468	4	US-09-252-991A-15339	Sequence 15339, A
c 158	35	3.5	474	4	US-09-252-991A-9	Sequence 9, Appli
c 159	35	3.5	882	3	US-08-818-112-138	Sequence 138, App
c 160	35	3.5	882	4	US-08-818-111-133	Sequence 133, App
c 161	35	3.5	882	4	US-09-056-556-138	Sequence 138, App
c 162	35	3.5	882	4	US-09-072-596-133	Sequence 133, App
c 163	35	3.5	1389	4	US-09-252-991A-15246	Sequence 15246, A
c 164	35	3.5	2124	4	US-09-266-965-44	Sequence 44, Appl
165	35	3.5	53500	4	US-09-266-965-76	Sequence 76, Appl
c 166	34.8	3.5	495	4	US-09-252-991A-3595	Sequence 3595, Ap
167	34.8	3.5	825	4	US-09-252-991A-3818	Sequence 3818, Ap
c 168	34.8	3.5	855	4	US-09-252-991A-3642	Sequence 3642, Ap
169	34.8	3.5	1224	4	US-09-252-991A-3865	Sequence 3865, Ap
170	34.8	3.5	1470	4	US-09-620-312D-122	Sequence 122, App
171	34.8	3.5	1785	4	US-09-252-991A-3774	Sequence 3774, Ap
c 172	34.8	3.5	1869	4	US-09-252-991A-3694	Sequence 3694, Ap
173	34.8	3.5	75395	4	US-09-984-890-3	Sequence 3, Appli
c 174	34.6	3.5	494	2	US-08-332-766A-22	Sequence 22, Appl
175	34.6	3.5	807	4	US-09-252-991A-14708	Sequence 14708, A
c 176	34.6	3.5	1041	2	US-08-580-545B-5	Sequence 5, Appli
c 177	34.6	3.5	1041	3	US-09-262-653A-5	Sequence 5, Appli
178	34.6	3.5	1965	4	US-09-205-258-116	Sequence 116, App
c 179	34.6	3.5	4177	3	US-09-023-082A-23	Sequence 23, Appl
c 180	34.6	3.5	5121	4	US-09-252-991A-15189	Sequence 15189, A
c 181	34.6	3.5	6327	4	US-09-252-991A-14931	Sequence 14931, A
c 182	34.4	3.4	421	2	US-08-332-766A-25	Sequence 25, Appl
c 183	34.4	3.4	3336	6	5457037-1	Patent No. 5457037
c 184	34.4	3.4	3702	4	US-09-252-991A-6666	Sequence 6666, Ap
185	34.4	3.4	18000	4	US-09-657-346A-17	Sequence 17, Appl
c 186	34.4	3.4	24183	3	US-08-943-731-3	Sequence 3, Appli
c 187	34.4	3.4	50000	4	US-09-146-053-3	Sequence 3, Appli
c 188	34.2	3.4	305	1	US-08-253-155A-4	Sequence 4, Appli
c 189	34.2	3.4	1563	4	US-09-252-991A-12880	Sequence 12880, A
c 190	34.2	3.4	1623	4	US-09-252-991A-12572	Sequence 12572, A
191	34.2	3.4	2532	4	US-09-252-991A-12912	Sequence 12912, A
192	34.2	3.4	87350	3	US-08-781-891-79	Sequence 79, Appl
193	34.2	3.4	87350	4	US-09-618-166-79	Sequence 79, Appl
194	34.2	3.4	87543	4	US-09-791-211-3	Sequence 3, Appli
c 195	34	3.4	981	4	US-09-252-991A-10487	Sequence 10487, A
196	34	3.4	1053	4	US-09-252-991A-10684	Sequence 10684, A
197	34	3.4	1549	4	US-09-444-336-5	Sequence 5, Appli
198	34	3.4	1920	4	US-09-620-312D-7	Sequence 7, Appli
199	34	3.4	50937	3	US-09-428-517-1	Sequence 1, Appli
c 200	34	3.4	64467	4	US-09-803-671B-3	Sequence 3, Appli
c 201	33.8	3.4	325	2	US-08-332-766A-11	Sequence 11, Appl
c 202	33.8	3.4	1830121	4	US-09-557-884-1	Sequence 1, Appli
c 203	33.8	3.4	1830121	4	US-09-643-990A-1	Sequence 1, Appli
c 204	33.6	3.4	1277	4	US-08-722-570-15	Sequence 15, Appl
c 205	33.6	3.4	1277	4	US-08-932-411A-16	Sequence 16, Appl
206	33.6	3.4	4521	4	US-09-533-494A-18	Sequence 18, Appl
c 207	33.4	3.3	334	2	US-08-623-906A-9	Sequence 9, Appli
208	33.4	3.3	846	4	US-09-252-991A-12628	Sequence 12628, A
c 209	33.4	3.3	1224	4	US-09-252-991A-13010	Sequence 13010, A
210	33.4	3.3	1293	4	US-09-252-991A-12462	Sequence 12462, A
c 211	33.4	3.3	1997	2	US-08-969-630-1	Sequence 1, Appli
212	33.4	3.3	2205	4	US-09-252-991A-16465	Sequence 16465, A
c 213	33.4	3.3	2256	4	US-09-252-991A-16137	Sequence 16137, A
214	33.2	3.3	180	4	US-09-841-334A-1	Sequence 1, Appli
215	33	3.3	367	2	US-08-332-766A-9	Sequence 9, Appli
c 216	33	3.3	370	2	US-08-332-766A-8	Sequence 8, Appli
217	33	3.3	674	3	US-08-871-809A-24	Sequence 24, Appl
c 218	33	3.3	885	4	US-09-252-991A-7532	Sequence 7532, Ap
219	33	3.3	912	4	US-09-252-991A-7811	Sequence 7811, Ap
220	33	3.3	1242	4	US-09-252-991A-7949	Sequence 7949, Ap
221	33	3.3	1740	4	US-09-252-991A-15553	Sequence 15553, A
222	33	3.3	1907	4	US-09-205-258-108	Sequence 108, App
223	33	3.3	246240	2	US-08-724-394A-20	Sequence 20, Appl
224	33	3.3	246240	2	US-08-724-394A-21	Sequence 21, Appl
225	33	3.3	246240	2	US-08-724-394A-22	Sequence 22, Appl
c 226	32.8	3.3	233	4	US-09-470-191-68	Sequence 68, Appl
227	32.8	3.3	283	4	US-09-313-294A-6805	Sequence 6805, Ap
c 228	32.8	3.3	329	1	US-08-253-155A-23	Sequence 23, Appl
229	32.8	3.3	799	3	US-08-998-416-420	Sequence 420, App
c 230	32.8	3.3	818	3	US-08-998-416-359	Sequence 359, App
c 231	32.8	3.3	909	4	US-09-252-991A-11391	Sequence 11391, A
232	32.8	3.3	948	4	US-09-252-991A-11445	Sequence 11445, A
233	32.8	3.3	1104	4	US-09-252-991A-12718	Sequence 12718, A
c 234	32.8	3.3	1743	4	US-09-252-991A-12913	Sequence 12913, A
c 235	32.6	3.3	819	4	US-09-252-991A-1742	Sequence 1742, Ap
c 236	32.6	3.3	822	4	US-09-252-991A-8940	Sequence 8940, Ap
c 237	32.6	3.3	1110	4	US-09-252-991A-8823	Sequence 8823, Ap
238	32.6	3.3	1734	4	US-09-252-991A-9061	Sequence 9061, Ap
239	32.6	3.3	2154	4	US-09-252-991A-9144	Sequence 9144, Ap
c 240	32.6	3.3	2397	1	US-07-891-942G-11	Sequence 11, Appl
241	32.6	3.3	10642	4	US-09-934-551-3	Sequence 3, Appli
c 242	32.4	3.2	485	2	US-08-332-766A-18	Sequence 18, Appl
c 243	32.4	3.2	624	4	US-09-252-991A-6551	Sequence 6551, Ap
244	32.4	3.2	765	4	US-09-252-991A-6511	Sequence 6511, Ap
245	32.4	3.2	864	4	US-09-252-991A-6520	Sequence 6520, Ap
246	32.4	3.2	2126	2	US-08-789-354-1	Sequence 1, Appli

	247	32.4	3.2	2126	3	US-09-110-937-1 Sequence 1, Appli
	248	32.4	3.2	2126	3	US-09-058-725B-1 Sequence 1, Appli
	249	32.4	3.2	2126	3	US-09-232-857-1 Sequence 1, Appli
c	250	32.4	3.2	6755	3	US-08-931-999-4 Sequence 4, Appli
c	251	32.2	3.2	286	4	US-09-313-294A-4886 Sequence 4886, Ap
c	252	32.2	3.2	287	4	US-09-313-294A-4106 Sequence 4106, Ap
c	253	32.2	3.2	759	4	US-09-252-991A-15651 Sequence 15651, A
	254	32.2	3.2	900	4	US-09-252-991A-1017 Sequence 1017, Ap
c	255	32.2	3.2	1458	4	US-09-252-991A-1112 Sequence 1112, Ap
	256	32.2	3.2	1461	4	US-09-252-991A-972 Sequence 972, App
c	257	32.2	3.2	1863	4	US-09-252-991A-733 Sequence 733, App
c	258	32.2	3.2	3010	4	US-09-996-243-313 Sequence 313, App
	259	32.2	3.2	4563	4	US-09-252-991A-930 Sequence 930, App
	260	32.2	3.2	33529	3	US-09-144-085-3 Sequence 3, Appli
	261	32.2	3.2	77536	4	US-09-410-551B-1 Sequence 1, Appli
c	262	32.2	3.2	246240	2	US-08-724-394A-20 Sequence 20, Appl
c	263	32.2	3.2	246240	2	US-08-724-394A-21 Sequence 21, Appl
c	264	32.2	3.2	246240	2	US-08-724-394A-22 Sequence 22, Appl
	265	32	3.2	320	3	US-09-165-264-13 Sequence 13, Appl
	266	32	3.2	373	3	US-08-871-809A-25 Sequence 25, Appl
	267	32	3.2	381	4	US-09-329-884-9 Sequence 9, Appli
c	268	32	3.2	471	4	US-09-252-991A-12181 Sequence 12181, A
	269	32	3.2	531	4	US-09-252-991A-12413 Sequence 12413, A
	270	32	3.2	567	4	US-09-252-991A-12364 Sequence 12364, A
	271	32	3.2	672	4	US-09-252-991A-3606 Sequence 3606, Ap
	272	32	3.2	690	4	US-09-252-991A-3701 Sequence 3701, Ap
c	273	32	3.2	747	4	US-09-252-991A-12243 Sequence 12243, A
c	274	32	3.2	751	4	US-09-227-357-117 Sequence 117, App
	275	32	3.2	822	4	US-09-252-991A-9029 Sequence 9029, Ap
c	276	32	3.2	1527	4	US-09-252-991A-8978 Sequence 8978, Ap
c	277	32	3.2	1551	4	US-09-252-991A-5309 Sequence 5309, Ap
c	278	32	3.2	1563	4	US-09-252-991A-15934 Sequence 15934, A
	279	32	3.2	1882	3	US-09-370-253-1 Sequence 1, Appli
	280	32	3.2	1953	4	US-09-252-991A-15763 Sequence 15763, A
	281	32	3.2	2019	4	US-09-252-991A-658 Sequence 658, App
c	282	32	3.2	2043	4	US-09-252-991A-8759 Sequence 8759, Ap
c	283	32	3.2	2052	4	US-09-252-991A-12103 Sequence 12103, A
c	284	32	3.2	3351	4	US-09-252-991A-15871 Sequence 15871, A
c	285	32	3.2	4158	4	US-09-252-991A-5348 Sequence 5348, Ap
	286	32	3.2	4953	4	US-09-252-991A-5227 Sequence 5227, Ap
	287	32	3.2	12047	2	US-09-022-461-1 Sequence 1, Appli
	288	32	3.2	12047	4	US-09-033-556-3 Sequence 3, Appli
	289	32	3.2	12047	4	US-09-474-699-11 Sequence 11, Appl
c	290	31.8	3.2	523	3	US-09-286-132-8 Sequence 8, Appli
c	291	31.8	3.2	523	4	US-09-836-941-8 Sequence 8, Appli
c	292	31.8	3.2	570	3	US-09-193-792-4 Sequence 4, Appli
c	293	31.8	3.2	597	3	US-09-286-132-7 Sequence 7, Appli
c	294	31.8	3.2	597	4	US-09-836-941-7 Sequence 7, Appli
c	295	31.8	3.2	882	4	US-09-252-991A-16022 Sequence 16022, A
c	296	31.8	3.2	1036	4	US-09-072-596-304 Sequence 304, App
c	297	31.8	3.2	1312	3	US-09-193-792-20 Sequence 20, Appl
	298	31.8	3.2	1456	2	US-08-951-148-6 Sequence 6, Appli
	299	31.8	3.2	1456	2	US-09-165-234-6 Sequence 6, Appli
	300	31.8	3.2	1456	3	US-09-274-570-6 Sequence 6, Appli

ALIGNMENTS

```
RESULT 1
US-08-770-379-17/c
; Sequence 17, Application US/08770379
; Patent No. 5849564
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
;    TITLE OF INVENTION:  HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,379
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  52342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  35100 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-770-379-17

  Query Match             12.8%;  Score 127.6;  DB 2;  Length 35100;
  Best Local Similarity   66.9%;  Pred. No. 7.6e-25;
  Matches  198;  Conservative    0;  Mismatches   94;  Indels    4;  Gaps    1;

Qy        465 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 524
              | ||  |||  || |    ||||| ||||| ||||    ||||| ||||| ||||    |
Db      24611 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24552

Qy        525 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 584
              |||| ||||| ||||    ||||| ||||| ||||    |||||||||| ||||    |
Db      24551 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24492

Qy        585 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 644
              |||| ||||| ||||    |||||||||| ||||    ||||| ||||| ||||    |
Db      24491 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24432

Qy        645 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT----GCCTGCTGGCTT 700
              |||| ||||| ||||    ||||| ||||| ||||    |||||    |  ||||| |
Db      24431 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24372

Qy        701 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTG 756
              |  | || |  ||||    ||||| || |  |||| |  ||||| ||| |  ||||
Db      24371 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTG 24316


RESULT 2
US-08-757-669A-17/c
; Sequence 17, Application US/08757669A
; Patent No. 6183751
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
;    TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
```

```
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  10036
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/757,669A
;       FILING DATE:
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  White, John P.
;       REGISTRATION NUMBER:  28,678
;       REFERENCE/DOCKET NUMBER:  45185-F
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 278-0400
;       TELEFAX:  (212) 391-0525
;   INFORMATION FOR SEQ ID NO:  17:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  35100 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-757-669A-17

  Query Match             12.8%;  Score 127.6;  DB 3;  Length 35100;
  Best Local Similarity   66.9%;  Pred. No. 7.6e-25;
  Matches  198;  Conservative    0;  Mismatches   94;  Indels    4;  Gaps    1;

Qy        465 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 524
              | ||  |||  || |    ||||| ||||| ||||    ||||| ||||| ||||    |
Db      24611 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24552

Qy        525 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 584
              |||| ||||| ||||    ||||| ||||| ||||    |||||||||| ||||    |
Db      24551 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24492

Qy        585 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 644
              |||| ||||| ||||    |||||||||| ||||    ||||| ||||| ||||    |
Db      24491 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24432

Qy        645 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT----GCCTGCTGGCTT 700
              |||| ||||| ||||    ||||| ||||| ||||    |||||    |  ||||| |
Db      24431 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24372

Qy        701 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTG 756
              |  | || |  ||||   |||||  || |  |||| |  ||||| ||| |  ||||
Db      24371 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTG 24316


RESULT 3
US-09-230-371A-17/c
; Sequence 17, Application US/09230371A
; Patent No. 6348586
; GENERAL INFORMATION:
;  APPLICANT:  Chang, Yuan
;  APPLICANT:  Bohenzky, Roy A
;  APPLICANT:  Russo, James J
;  APPLICANT:  Edelman, Isidore S
;  APPLICANT:  Moore, Patrick S
;  TITLE OF INVENTION:  UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE:  45185-G-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/230,371A
;  CURRENT FILING DATE:  1999-11-17
;  PRIOR APPLICATION NUMBER:  PCT/US97/13346
;  PRIOR FILING DATE:  1997-07-22
;  NUMBER OF SEQ ID NOS:  30
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 17
;   LENGTH: 35100
;   TYPE: DNA
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-17

  Query Match             12.8%;  Score 127.6;  DB 4;  Length 35100;
  Best Local Similarity   66.9%;  Pred. No. 7.6e-25;
  Matches  198;  Conservative    0;  Mismatches   94;  Indels    4;  Gaps    1;

Qy        465 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 524
              | ||  |||  || |    ||||| ||||| ||||    ||||| ||||| ||||    |
Db      24611 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24552

Qy        525 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 584
              |||| ||||| ||||    ||||| ||||| ||||    |||||||||| ||||    |
Db      24551 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24492

Qy        585 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 644
              |||| ||||| ||||    |||||||||| ||||    ||||| ||||| ||||    |
Db      24491 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24432

Qy        645 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT----GCCTGCTGGCTT 700
              |||| ||||| ||||    ||||| ||||| ||||    |||||    |  ||||| |
Db      24431 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24372

Qy        701 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTG 756
              |  | || |  ||||   |||||  || |  |||| |  ||||| ||| |  ||||
Db      24371 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTG 24316


RESULT 4
PCT-US93-06251-23/c
; Sequence 23, Application PC/TUS9306251
;  GENERAL INFORMATION:
;    APPLICANT:  Wickstrom, Eric and Rife, Jason P.
;    TITLE OF INVENTION:  Trivalent Synthesis of Oligonucleotides Containing
;    TITLE OF INVENTION:  Stereospecific Alkylphosphonates and Arylphosphonates
;    NUMBER OF SEQUENCES:  93
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCULLY, SCOTT, MURPHY & PRESSER
;      STREET:  400 Garden City Plaza
;      CITY:  Garden City
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  11530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/06251
;      FILING DATE:  19930630
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DiGiglio, Frank S.
;      REGISTRATION NUMBER:  31,346
;      REFERENCE/DOCKET NUMBER:  8586
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  516-742-4343
;      TELEFAX:  516-742-4366
;      TELEX:  230 901 SANS UR
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11558 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
PCT-US93-06251-23

  Query Match             10.2%;  Score 101.8;  DB 5;  Length 11558;
  Best Local Similarity   47.9%;  Pred. No. 4.7e-18;
```

```
  Matches  326;  Conservative    0;  Mismatches  352;  Indels    3;  Gaps    1;

Qy        145 GACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 204
              | |||  ||   | | | ||   |||  |  ||   | | | | | |||  || | | ||
Db       6365 GTCAGATGGGTGGGTGGATGGATGGATGGGTGGATGGATGGATGGATTGATGGATG 6306

Qy        205 AGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTC 264
                |||  |  |||  | ||| | | | |  ||   | | | ||||  | | | | | |
Db       6305 GATGGATGGATGGGTGGATGGGTGGATGGATGGATGGGTGGATGGCTGGATGGATGGATG 6246

Qy        265 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 324
              | |||   | | | | | | | |||   | | | | | | | ||   | | | | |||
Db       6245 GGTGGATGGATGGATAGGTGGATGGATGGATAGGTGGATGGATGGATGGGTGGATGGCTG 6186

Qy        325 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 384
              | |||   | | | | | | | |||   | | | | | | | |||   | | | | | |
Db       6185 GATGGATGGATGGGTGGATGGGTGGATGGATGGATGGATGGATGGATGGATGGATGGATG 6126

Qy        385 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 444
              | |||   | | | | | | | |||   ||| | ||   |||  |  ||   | | | |
Db       6125 GATGGGTGGATGGGTGGATGGGTGGATGGCTGGTTGGCTGGATGGATGGATGGGTGGATG 6066

Qy        445 GCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAG 504
              | |  | ||  ||   |   |  |||| | ||| |  ||   | |||   ||| || |
Db       6065 GGTGGGTGGGTGGATGGATGGATGGATGGATGGATGGATGGGTGGATGGCTGGTTG 6006

Qy        505 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 564
              |   | ||||  | | |  ||   | ||| | | |  |||  | |||   | | |  |
Db       6005 GATGGATGGGTGGATGGATGGATAGGTGGACGGATGGATGGGTGGATGGCTGGATGGATG 5946

Qy        565 GGCTCCTGGGCTGCTAGTAGGGCTGCTG---GGCTGCTAGTAGGGCTCCTGGGCTGCTAG 621
              ||    | |   | | |   ||  | ||   || ||   | | || |    || ||  ||
Db       5945 GGTGGATTGCTGGATGGATTGGTGGATGGGTGGATGGGTGGATGGGTGGATGGGTGAAAG 5886

Qy        622 TAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAG 681
              | ||  |   ||  |       || | | ||| |   |    |||| | ||| | | |
Db       5885 GATGGAAGGAAGGAGGGAGCATGGACGGATGGACATGTGAATGGGCAGATGGACAGATGA 5826

Qy        682 TAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGG 741
              ||   |   | ||| | |   |  | |         ||   |   |   |       |
Db       5825 ACGGTTGGAAGGATGGATGGAAAGAAGGATGAGAGAAAGGAAAGAAGGAAGGAAGGAGAG 5766

Qy        742 GCTGCTGTGCAGCTGGGAGAACAGAGTAGGGCTGCCGGCCAGCTGCGTGCGAGGGCGTCC 801
              |    ||  ||| |||| ||| ||| ||  |     ||  ||  |   |  ||   |
Db       5765 GTGTATGGACAGATGGGTGAAGAGAATAAAGGAAGGGGAAAGAAGGAAGGAAGATGGATG 5706

Qy        802 GAGGGCCAGACGAGGACACGG 822
              || |||  ||| |  | | ||
Db       5705 GATGGCTTGACAAACAGATGG 5685


RESULT 5
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
;   GENERAL INFORMATION:
;        APPLICANT: Fleischmann et al.
;        TITLE OF INVENTION: The Nucleotide sequence of
;                            the Haemophilus influenzae Rd Genome, Fragments
;                            Thereof, and Uses Thereof
;        NUMBER OF SEQUENCES: 1
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: MD
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3 1/2 inch diskette
;             COMPUTER: Dell Pentium
;             OPERATING SYSTEM: MS DOS v6.22
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/557,884
;             FILING DATE: 25-Apr-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/476,102
;             FILING DATE: JUN-5-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Michelle S. Marks
;             REGISTRATION NUMBER: 41,971
;             REFERENCE/DOCKET NUMBER: PB186P3
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 301-309-8504
;             TELEFAX: 301-309-8439
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1830121 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

  Query Match             7.7%;  Score 77.4;  DB 4;  Length 1830121;
  Best Local Similarity   84.5%;  Pred. No. 1.2e-10;
  Matches   87;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy        319 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 378
              ||||| ||| || |||||||||||| ||| || |||||||||||| ||| || |||||||
Db    1543129 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 1543188

Qy        379 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGA 421
              ||||| ||| || |||||||||||| ||| || ||||||| ||
Db    1543189 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCAGA 1543231


RESULT 6
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
;   GENERAL INFORMATION:
;        APPLICANT: Robert D. Fleischmann
;                   Mark D. Adams
;                   Owen White
;                   Hamilton O. Smith
;                   J. Craig Venter
;        TITLE OF INVENTION: The Nucleotide sequence of
;                            the Haemophilus influenzae Rd Genome, Fragments
;                            Thereof, and Uses Thereof
;        NUMBER OF SEQUENCES: 1
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville,
;             STATE: MD
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3 1/2 inch diskette
;             COMPUTER: Dell Pentium
;             OPERATING SYSTEM: MS DOS v6.22
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/643,990A
;             FILING DATE: 23-Aug-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/487,429
;             FILING DATE: 1995-06-07
;             APPLICATION NUMBER: 08/426,787
;             FILING DATE: 1995-04-21
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Kenley K. Hoover
```

```
;             REGISTRATION NUMBER: 40,302
;             REFERENCE/DOCKET NUMBER: PB186P1C1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 301-610-5790
;             TELEFAX: 310-309-8439
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1830121 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

  Query Match             7.7%;  Score 77.4;  DB 4;  Length 1830121;
  Best Local Similarity   84.5%;  Pred. No. 1.2e-10;
  Matches   87;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy        319 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 378
              ||||| ||| || |||||||||||| ||| || |||||||||||| ||| || |||||||
Db    1543129 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 1543188

Qy        379 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGA 421
              ||||| ||| || |||||||||||| ||| || ||||||| ||
Db    1543189 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCAGA 1543231


RESULT 7
US-08-533-669A-7/c
; Sequence 7, Application US/08533669A
; Patent No. 5834592
;  GENERAL INFORMATION:
;    APPLICANT:  Corixa Corporation
;    TITLE OF INVENTION:  LEISHMANIA ANTIGENS FOR USE IN THE
;    TITLE OF INVENTION:  THERAPY AND DIAGNOSIS OF LEISHMANIASIS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/533,669A
;      FILING DATE:  22-SEP-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1771 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1698
US-08-533-669A-7

  Query Match             6.2%;  Score 61.8;  DB 2;  Length 1771;
  Best Local Similarity   48.1%;  Pred. No. 1.5e-07;
  Matches  295;  Conservative    0;  Mismatches  312;  Indels    6;  Gaps    4;

Qy        137 GCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 196
              |||  | | | ||  |  |||| | ||   ||| | |   ||  || ||||||    | |
Db        961 GCTCCGCGCAGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGT 902

Qy        197 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 256
              || ||       | | |   | || || |   ||  | |     ||  ||| ||| | |
Db        901 GCATGGCCTCCAGGCGGGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGT 842

Qy        257 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 316
              || ||    |  | | |   || | ||||   |||| |      || ||  | |  || |
Db        841 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGC-T 783

Qy        317 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 376
              || | ||||  ||||||   ||  ||  ||| | |  | ||| ||     |  | | |
Db        782 GCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCC 723

Qy        377 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC---CTGCTTGCTGAGGGGACAGTAGG 433
              || | ||||   |||| |      || ||  | |   |||| | ||   ||| | | | |
Db        722 GCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGACG 663

Qy        434 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 493
                 || |||| |     | |   |    || | |     || |  | |  || |||
Db        662 CGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACA 603

Qy        494 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 553
                 ||  |   |||  |  |  ||  |    ||| |   | || || |   | || | |
Db        602 CGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGC 543

Qy        554 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 613
                | ||  | | | |  | || || |  || |||   |   ||  |   ||| |  |
Db        542 TCCGCGCGGA-GACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTG 484

Qy        614 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 673
                  ||       ||| |   | || || |     || | |  | || |  | ||    ||| |
Db        483 CATGGCCTCCAGGCGGGCCGCCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-G 425

Qy        674 GCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTG 733
                 || || |  || |||   || ||| |||| |   || |||  |     | |   | | |
Db        424 CCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCG 365

Qy        734 CTACTAGGGCTGC 746
              |  |   |||| |
Db        364 CCTCCTCGGCTTC 352


RESULT 8
US-08-511-872-1/c
; Sequence 1, Application US/08511872
; Patent No. 5965142
;  GENERAL INFORMATION:
;    APPLICANT:  Dillon, Davin C.
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Day, Craig H.
;    TITLE OF INVENTION:  POLYPEPTIDES AND METHODS FOR THE
;    TITLE OF INVENTION:  DETECTION OF L. tropica INFECTION
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  94104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/511,872
;      FILING DATE:  04-AUG-1995
;      CLASSIFICATION:  424
```

```
;   , ATTORNEY/AGENT INFORMATION:
;       NAME:  No. 5965142tenburg, Carol
;       REGISTRATION NUMBER:  39,317
;       REFERENCE/DOCKET NUMBER:  210121.405
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;       TELEX:  3723836 SEEDANDBERRY
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1771 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..1698
US-08-511-872-1

  Query Match             6.2%;  Score 61.8;  DB 2;  Length 1771;
  Best Local Similarity   48.1%;  Pred. No. 1.5e-07;
  Matches  295;  Conservative    0;  Mismatches  312;  Indels    6;  Gaps    4;

Qy        137 GCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 196
              |||  | | | ||  |  |||| | ||   ||| | |   ||  || ||||||    | |
Db        961 GCTCCGCGCAGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGT 902

Qy        197 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 256
              || ||       | | |   | || || |   ||  | |     ||  ||| |||  | |
Db        901 GCATGGCCTCCAGGCGGGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGT 842

Qy        257 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 316
              || ||    |  | | |   || | ||||   |||| |      || ||  | |  || |
Db        841 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGC-T 783

Qy        317 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 376
              || | ||||  ||||||   ||  ||  ||| | |  | ||| ||   | | | |
Db        782 GCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCC 723

Qy        377 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC---CTGCTTGCTGAGGGGACAGTAGG 433
              || | ||||   |||| |      || ||  | |   |||| | ||   ||| | | | |
Db        722 GCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGACG 663

Qy        434 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 493
                || |||| |     | |   |    || | |     || |  | |  || |||
Db        662 CGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACA 603

Qy        494 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 553
                ||  |   |||  |  |  ||  |   ||| |   | || || |   | || | |
Db        602 CGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGC 543

Qy        554 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 613
              | ||  | | | | |   | || || |  || |||   |   ||  |   |||  |  |
Db        542 TCCGCGCGGA-GACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTG 484

Qy        614 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 673
                ||       ||| |   | || || |     || | |  | ||  | ||    ||| |
Db        483 CATGGCCTCCAGGCGGGCCGCCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-G 425

Qy        674 GCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTG 733
              || || |  || |||   || ||| |||| |   || |||  |     | |   | | |
Db        424 CCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCG 365

Qy        734 CTACTAGGGCTGC 746
              |  |   |||| |
Db        364 CCTCCTCGGCTTC 352


RESULT 9
US-09-183-861-7/c
; Sequence 7, Application US/09183861
; Patent No. 6365165
;  GENERAL INFORMATION:
;     APPLICANT:  Reed, Steven G.
;     APPLICANT:  Campos-Neto, Antonio
;     APPLICANT:  Webb, John R.
;     APPLICANT:  Dillon, Davin C.
;     APPLICANT:  Skeiky, Yasir A.W.
;     TITLE OF INVENTION:  LEISHMANIA ANTIGENS FOR USEIN THE THERAPY AND
;     NUMBER OF SEQUENCES:  87
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SEED and BERRY LLP
;       STREET:  6300 Columbia Center, 701 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98104-7092
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/183,861
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  09/022,765
;       FILING DATE:  12-FEB-1998
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Maki, David J.
;       REGISTRATION NUMBER:  31,392
;       REFERENCE/DOCKET NUMBER:  210121.420C3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1771 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..1698
US-09-183-861-7

  Query Match             6.2%;  Score 61.8;  DB 4;  Length 1771;
  Best Local Similarity   48.1%;  Pred. No. 1.5e-07;
  Matches  295;  Conservative    0;  Mismatches  312;  Indels    6;  Gaps    4;

Qy        137 GCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 196
              |||  | | | ||  |  |||| | ||   ||| | |   ||  || ||||||    | |
Db        961 GCTCCGCGCAGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGT 902

Qy        197 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 256
              || ||       | | |   | || || |   ||  | |     ||  ||| |||  | |
Db        901 GCATGGCCTCCAGGCGGGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGT 842

Qy        257 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 316
              || ||    |  | | |   || | ||||   |||| |      || ||  | |  || |
Db        841 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGC-T 783

Qy        317 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 376
              || | ||||  ||||||   ||  ||  ||| | |  | ||| ||   | | | |
Db        782 GCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCC 723

Qy        377 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC---CTGCTTGCTGAGGGGACAGTAGG 433
              || | ||||   |||| |      || ||  | |   |||| | ||   ||| | | | |
Db        722 GCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGACG 663

Qy        434 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 493
                || |||| |     | |   |    || | |     || |  | |  || |||
Db        662 CGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACA 603

Qy        494 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 553
```

```
Db        602 CGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGC 543

Qy        554 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 613

Db        542 TCCGCGCGGA-GACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTG 484

Qy        614 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 673

Db        483 CATGGCCTCCAGGCGGGCCGCCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-G 425

Qy        674 GCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTG 733

Db        424 CCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCG 365

Qy        734 CTACTAGGGCTGC 746

Db        364 CCTCCTCGGCTTC 352


RESULT 10
US-09-022-765-7/c
; Sequence 7, Application US/09022765
; Patent No. 6375955
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Campos-Neto, Antonio
;    APPLICANT:  Webb, John R.
;    APPLICANT:  Dillon, Davin C.
;    APPLICANT:  Skeiky, Yasir A.W.
;    TITLE OF INVENTION:  LEISHMANIA ANTIGENS FOR USEIN THE THERAPY AND
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/022,765
;      FILING DATE:  12-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.420C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1771 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1698
US-09-022-765-7

  Query Match             6.2%;  Score 61.8;  DB 4;  Length 1771;
  Best Local Similarity   48.1%;  Pred. No. 1.5e-07;
  Matches  295;  Conservative    0;  Mismatches  312;  Indels    6;  Gaps    4;

Qy        137 GCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 196

Db        961 GCTCCGCGCAGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGT 902

Qy        197 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 256

Db        901 GCATGGCCTCCAGGCGGGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGT 842

Qy        257 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 316

Db        841 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGC-T 783

Qy        317 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 376

Db        782 GCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCC 723

Qy        377 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC---CTGCTTGCTGAGGGGACAGTAGG 433

Db        722 GCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGACG 663

Qy        434 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 493

Db        662 CGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACA 603

Qy        494 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 553

Db        602 CGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGC 543

Qy        554 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 613

Db        542 TCCGCGCGGA-GACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTG 484

Qy        614 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 673

Db        483 CATGGCCTCCAGGCGGGCCGCCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-G 425

Qy        674 GCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTG 733

Db        424 CCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCG 365

Qy        734 CTACTAGGGCTGC 746

Db        364 CCTCCTCGGCTTC 352


RESULT 11
US-09-551-974A-7/c
; Sequence 7, Application US/09551974A
; Patent No. 6500437
; GENERAL INFORMATION:
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Campos-Neto, Antonio
;  APPLICANT:  Webb, John R.
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Skeiky, Yasir A.W.
;  TITLE OF INVENTION:  LEISHMANIA ANTIGENS FOR USE IN THE
;  TITLE OF INVENTION:  THERAPY AND DIAGNOSIS OF LEISHMANIASIS
;  FILE REFERENCE:  210121.420C5
;  CURRENT APPLICATION NUMBER: US/09/551,974A
;  CURRENT FILING DATE:  2000-04-14
;  NUMBER OF SEQ ID NOS:  101
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 1771
;   TYPE: DNA
;   ORGANISM: Leishmania tropica
US-09-551-974A-7

  Query Match             6.2%;  Score 61.8;  DB 4;  Length 1771;
  Best Local Similarity   48.1%;  Pred. No. 1.5e-07;
  Matches  295;  Conservative    0;  Mismatches  312;  Indels    6;  Gaps    4;

Qy        137 GCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 196

Db        961 GCTCCGCGCAGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGT 902
```

```
Qy        197 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 256
              || ||        | | |   | || || |   ||  | |      ||  ||| |||  | |
Db        901 GCATGGCCTCCAGGCGGGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGT 842

Qy        257 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 316
              || ||    |  | | |   || | ||||   |||| |       || ||  | |  || |
Db        841 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGC-T 783

Qy        317 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 376
              || | ||||/ ||||||   ||  ||  ||| | |  | ||| ||    |  | | |
Db        782 GCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCC 723

Qy        377 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC---CTGCTTGCTGAGGGGACAGTAGG 433
              || | ||||   |||| |       || ||  | |   |||| | ||   ||| | | | |
Db        722 GCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGACG 663

Qy        434 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 493
              || |||| |     | |   |    || | |     || |  | |  || |||
Db        662 CGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACA 603

Qy        494 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 553
              ||  |   |||  |  |  ||  |    ||| |   | || || |   | || | |
Db        602 CGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGC 543

Qy        554 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 613
              | ||  | | | | |   | || || |  || |||   |   ||  |   |||  |  |
Db        542 TCCGCGCGGA-GACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTG 484

Qy        614 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 673
              ||       ||| |   | || || |     || | |  | ||  | ||    ||| |
Db        483 CATGGCCTCCAGGCGGGCCGCCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-G 425

Qy        674 GCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTG 733
              || || |  || |||   || ||| |||| |   || |||  |     | |   | | |
Db        424 CCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCG 365

Qy        734 CTACTAGGGCTGC 746
              |  |   |||| |
Db        364 CCTCCTCGGCTTC 352


RESULT 12
US-09-146-053-6/c
; Sequence 6, Application US/09146053A
; Patent No. 6399349
; GENERAL INFORMATION:
;  APPLICANT: Ryan, James W.
;  APPLICANT: Sprinkle, Terry Joe Curtis
;  APPLICANT: Venema, Richard C.
;  TITLE OF INVENTION: Human Aminopeptidase P Gene
;  FILE REFERENCE: MCG103
;  CURRENT APPLICATION NUMBER: US/09/146,053A
;  CURRENT FILING DATE: 1998-09-02
;  EARLIER APPLICATION NUMBER: 60/057,854
;  EARLIER FILING DATE: 1997-09-02
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 45546
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-146-053-6

  Query Match             6.1%;  Score 60.8;  DB 4;  Length 45546;
  Best Local Similarity   60.1%;  Pred. No. 9.4e-07;
  Matches  101;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db      18750 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTCT 18691

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
              | ||| ||  ||| ||  |   ||  ||| ||| ||  ||  ||| ||| ||  ||  |
Db      18690 TTCTTTCTCTCTCTCTCTCTCTCTCTCTTTCTCTCTCTCCCTCTTTCTTTCTTTCTTTCT 18631

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||
Db      18630 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTCCT 18583


RESULT 13
US-08-476-102A-4/c
; Sequence 4, Application US/08476102A
; Patent No. 6355450
;   GENERAL INFORMATION:
;        APPLICANT: Fleischmann, et al.
;        TITLE OF INVENTION: Nucleotide Sequence of the Haemphilus
; Patent No. 6355450
;                              influenzae Rd Genome, Fragments Thereof, and Uses Thereo
;        NUMBER OF SEQUENCES: 9
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Ave.
;             CITY: Rockville
;             STATE: MD
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/476,102A
;             FILING DATE: 07-Jun-1995
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Marks, Michelle S.
;             REGISTRATION NUMBER: 41,971
;             REFERENCE/DOCKET NUMBER: PB186P2
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 301-309-8504
;             TELEX: 301-309-8439
;   INFORMATION FOR SEQ ID NO: 4:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 3252 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-08-476-102A-4

  Query Match             5.8%;  Score 58;  DB 4;  Length 3252;
  Best Local Similarity   58.8%;  Pred. No. 2.1e-06;
  Matches  100;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy        275 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 334
              ||| ||| | | | |  || ||| ||| | | | |  || ||| ||| | | | |  ||
Db        222 TTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGG 163

Qy        335 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 394
              ||| ||| | | | |  || ||| ||| | | | |  || ||| ||| | | | |  ||
Db        162 TTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGG 103

Qy        395 TTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 444
              ||| ||| | | | |  || ||| |  ||   ||       |||||||
Db        102 TTGGTTGGTTGGTTGGTTGGTTGGTTGGGCTGCATAGACTGTGCTTGCAT 53


RESULT 14
PCT-US96-05320A-714/c
; Sequence 714, Application PC/TUS9605320A
;  GENERAL INFORMATION:
;    APPLICANT:  Human Genome Sciences
;    APPLICANT:  9410 Key West Avenue
```

```
;    APPLICANT:  Rockville, MD 20850
;    APPLICANT:  United States of America
;    APPLICANT:  Johns Hopkins University
;    APPLICANT:  720 Rutland Avenue
;    APPLICANT:  Baltimore, MD 21205
;    APPLICANT:  United States of America
;    APPLICANT:  Mark D. Adams
;    APPLICANT:  Owen White
;    APPLICANT:  Hamilton O. Smith
;    APPLICANT:  J. Craig Venter
;    TITLE OF INVENTION:  Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,
;    NUMBER OF SEQUENCES:  48
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20003-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/05320A
;      FILING DATE:  April22, 1996
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/476,102
;      FILING DATE:  June 7, 1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/487,429
;      FILING DATE:  June 7, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Eric K. Steffe
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.014PC01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  714:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3252 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
PCT-US96-05320A-714

  Query Match             5.8%;  Score 58;  DB 5;  Length 3252;
  Best Local Similarity   58.8%;  Pred. No. 2.1e-06;
  Matches  100;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy        275 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 334
              ||| ||| | | | |  || ||| ||| | | | |  || ||| ||| | | | |  ||
Db        222 TTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGG 163

Qy        335 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 394
              ||| ||| | | | |  || ||| ||| | | | |  || ||| ||| | | | |  ||
Db        162 TTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGG 103

Qy        395 TTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 444
              ||| ||| | | | |  || ||| |  ||   ||       |||||| |
Db        102 TTGGTTGGTTGGTTGGTTGGTTGGGCTGCATAGACTGTGCTTGCAT 53


RESULT 15
US-08-332-766A-20/c
; Sequence 20, Application US/08332766A
; Patent No. 5843647
;  GENERAL INFORMATION:
;    APPLICANT:  JEFFREYS, Alec J.
;    APPLICANT:  ARMOUR, John
;    TITLE OF INVENTION:  SIMPLE TANDEM REPEATS
;    NUMBER OF SEQUENCES:  125
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D. C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/332,766A
;      FILING DATE:  01-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9326052.9
;      FILING DATE:  21-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BIRD, Donald J.
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  217211/M94/0434/GB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  350 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-332-766A-20

  Query Match             5.7%;  Score 56.8;  DB 2;  Length 350;
  Best Local Similarity   54.2%;  Pred. No. 1.9e-06;
  Matches  135;  Conservative    0;  Mismatches  113;  Indels    1;  Gaps    1;

Qy        172 AGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTG-CTGGCTTGC 230
              ||   |||||| |||  ||  |||||| |||         |       ||| ||| ||| |
Db        325 AGCCTGCCTGCCTGCCTGCCTGCCTGCCTTCCTTCCTTNCTCCCTGACTGCCTTCC 266

Qy        231 TAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 290
              |        |  | ||  | | ||| ||| ||  ||  |   ||| ||| ||   |  |
Db        265 TTCCTTCCTTCTTTCCTTCCTTCCTTCCTTCCTACCTTCCTACCTTCCTTCCTTCTTTCC 206

Qy        291 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 350
              | ||| |||  |  ||  | | |||  || ||  ||  | | ||  ||  ||  ||| |
Db        205 TTCCTTCCTTCTTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCCTGCCTGCCTGCCTGCC 146

Qy        351 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 410
                ||| || ||  ||  ||| ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db        145 TGCCTTCTTTCCTTCCTTCCCTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCC 86

Qy        411 CTGCTTGCT 419
              | ||| ||
Db         85 TTCCTTCCT 77


RESULT 16
US-08-332-766A-24/c
; Sequence 24, Application US/08332766A
; Patent No. 5843647
;  GENERAL INFORMATION:
;    APPLICANT:  JEFFREYS, Alec J.
;    APPLICANT:  ARMOUR, John
;    TITLE OF INVENTION:  SIMPLE TANDEM REPEATS
;    NUMBER OF SEQUENCES:  125
;    CORRESPONDENCE ADDRESS:
```

```
;     ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;     STREET:  1100 New York Avenue, N.W.
;     CITY:  Washington
;     STATE:  D. C.
;     COUNTRY:  U.S.A.
;     ZIP:  20005-3918
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/332,766A
;     FILING DATE:  01-NOV-1994
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  GB 9326052.9
;     FILING DATE:  21-DEC-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  BIRD, Donald J.
;     REGISTRATION NUMBER:  25,323
;     REFERENCE/DOCKET NUMBER:  217211/M94/0434/GB
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (202) 861-3000
;     TELEFAX:  (202) 822-0944
;     TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  24:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  538 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  DNA (genomic)
US-08-332-766A-24

  Query Match             5.4%;  Score 54.4;  DB 2;  Length 538;
  Best Local Similarity   55.2%;  Pred. No. 1e-05;
  Matches  106;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        230 CTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 289
              |||||       ||| ||  |   ||  ||  ||| ||  |   ||  ||  ||| ||
Db        299 CTAGTCCTTTCCCTCCCTCCCTCCCTCCCTCCCTTCCTCCCTCCCTCCCTCCCTCC 240

Qy        290 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 349
              | | ||  ||| ||  ||  ||| ||| ||| |   ||  ||| ||| ||| |   ||
Db        239 CTTCCTCCCTTCCTTCCTTCCCTCCTTCCTTCCCTCCTTCCCTCCTTCCTTCCCTCCTTC 180

Qy        350 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 409
              | | | ||| ||  ||  ||| ||| ||| ||  ||  | | ||| ||| |   ||
Db        179 CTTCCCTCCTTCCTTCCTTCCCTCCTTCCTTCCTTCCTTCCATTCTTTCTTTCATTCTTT 120

Qy        410 CCTGCTTGCTGA 421
              | | |   || |
Db        119 CTTCCCCTCTTA 108


RESULT 17
US-09-268-992-7/c
; Sequence 7, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
;  APPLICANT: Chen, H.
;  APPLICANT: Freimer, N.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
;  TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
;  FILE REFERENCE: 7853-138
;  CURRENT APPLICATION NUMBER: US/09/268,992
;  CURRENT FILING DATE: 1999-03-16
;  EARLIER APPLICATION NUMBER: 09/236,134
;  EARLIER FILING DATE: 1999-01-22
;  EARLIER APPLICATION NUMBER: 60/106,056
;  EARLIER FILING DATE: 1998-10-28
;  EARLIER APPLICATION NUMBER: 60/088,312
;  EARLIER FILING DATE: 1998-06-05
;  EARLIER APPLICATION NUMBER: 60/078,044
;  EARLIER FILING DATE: 1998-03-16
;  NUMBER OF SEQ ID NOS: 84
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 72604
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: all n positions
;   OTHER INFORMATION: n=a, c, g, or t
US-09-268-992-7

  Query Match             5.4%;  Score 53.8;  DB 4;  Length 72604;
  Best Local Similarity   58.4%;  Pred. No. 8.7e-05;
  Matches   94;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        243 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 302
              ||  |  | | |||  || ||| ||  | | ||| ||| ||  ||  | | ||| ||| |
Db      45313 TCTTTCTCTTTCTTCTTTTCTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTYCTTYC 45254

Qy        303 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 362
              |  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| |
Db      45253 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC 45194

Qy        363 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403
              ||  |       ||  ||  ||| ||  | | ||| ||| ||
Db      45193 TCTTTCTTTCTCTCTCTCTCTCTTTCTTTCTTTCTTTCT 45153


RESULT 18
US-09-657-474-7/c
; Sequence 7, Application US/09657474
; Patent No. 6399762
; GENERAL INFORMATION:
;  APPLICANT: Chen, H.
;  APPLICANT:  Freimer, N.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
;  TITLE OF INVENTION:  AND TREATING CHROMOSOME-18p RELATED DISORDERS
;  FILE REFERENCE: 7853-138
;  CURRENT APPLICATION NUMBER: US/09/657,474
;  CURRENT FILING DATE:  2000-09-07
;  PRIOR APPLICATION NUMBER: 09/268,992
;  PRIOR FILING DATE: 1999-03-16
;  PRIOR APPLICATION NUMBER: 09/236,134
;  PRIOR FILING DATE: 1999-01-22
;  PRIOR APPLICATION NUMBER: 60/106,056
;  PRIOR FILING DATE: 1998-10-28
;  PRIOR APPLICATION NUMBER: 60/088,312
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/078,044
;  PRIOR FILING DATE: 1998-03-16
;  NUMBER OF SEQ ID NOS: 84
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 72604
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: all n positions
;   OTHER INFORMATION: n=a, c, g, or t
US-09-657-474-7

  Query Match             5.4%;  Score 53.8;  DB 4;  Length 72604;
  Best Local Similarity   58.4%;  Pred. No. 8.7e-05;
  Matches   94;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        243 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 302
              ||  |  | | |||  || ||| ||  | | ||| ||| ||  ||  | | ||| ||| |
Db      45313 TCTTTCTCTTTCTTCTTTTCTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTYCTTYC 45254
```

```
Qy        303 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 362
              |  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| |
Db      45253 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC 45194

Qy        363 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403
              ||  |      ||  ||  ||| ||  | | ||| ||| ||
Db      45193 TCTTTCTTTCTCTCTCTCTCTCTCTTTCTTTCTTTCTTTCT 45153


RESULT 19
US-09-539-333D-1/c
; Sequence 1, Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Chumakov, Ilya
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Bihain, Bernard
;  APPLICANT:  Essioux, Laurent
;  TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
;  FILE REFERENCE: GENSET.047AUS
;  CURRENT APPLICATION NUMBER: US/09/539,333D
;  CURRENT FILING DATE:  2000-03-30
;  PRIOR APPLICATION NUMBER: US 60/126,903
;  PRIOR FILING DATE: 1999-03-30
;  PRIOR APPLICATION NUMBER: US 60/131,971
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: US 60/132,065
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: US 60/143,928
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 60/145,915
;  PRIOR FILING DATE: 1999-07-27
;  PRIOR APPLICATION NUMBER: US 60/146,453
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 60/146,452
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 60/162,288
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: US 09/416,384
;  PRIOR FILING DATE: 1999-10-12
;  NUMBER OF SEQ ID NOS: 231
;  SOFTWARE: Patent.pm
; SEQ ID NO 1
;   LENGTH: 319608
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 31..1107
;   OTHER INFORMATION: 5'regulatory region g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 1108..1289
;   OTHER INFORMATION: exon A g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 14877..14920
;   OTHER INFORMATION: exon B g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 18778..18862
;   OTHER INFORMATION: exon Bbis g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 25593..25740
;   OTHER INFORMATION: exon C g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 29388..29502
;   OTHER INFORMATION: exon D g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 29967..30282
;   OTHER INFORMATION: exon E g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 64666..64812
;   OTHER INFORMATION: exon F g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 65505..65853
;   OTHER INFORMATION: exon G g35018 gene
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 65854..67854
;   OTHER INFORMATION: 3'regulatory region g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 94124..94964
;   OTHER INFORMATION: exon g35017
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 201188..201234
;   OTHER INFORMATION: exon S g35030 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 214676..214793
;   OTHER INFORMATION: exon T g35030 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 215702..215746
;   OTHER INFORMATION: exon U g35030 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 216836..216915
;   OTHER INFORMATION: exon V g35030 gene
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 213818..215818
;   OTHER INFORMATION: 3'regulatory region g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 215819..215941
;   OTHER INFORMATION: exon R complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 215819..215975
;   OTHER INFORMATION: exon Rbis complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 216661..216952
;   OTHER INFORMATION: exon Qbis complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 216661..217061
;   OTHER INFORMATION: exon Q complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 217027..217061
;   OTHER INFORMATION: exon Q1 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 229647..229742
;   OTHER INFORMATION: exon X complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 230408..230721
;   OTHER INFORMATION: exon P complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 231272..231412
;   OTHER INFORMATION: exon Obis complement g34872 gene
;   FEATURE:
```

```
;   NAME/KEY: exon
;   LOCATION: 231787..231880
;   OTHER INFORMATION: exon O2 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 231870..231879
;   OTHER INFORMATION: exon O1 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 234174..234321
;   OTHER INFORMATION: exon O complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 237406..237428
;   OTHER INFORMATION: exon Nbis complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 239719..239807
;   OTHER INFORMATION: exon N2 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 239719..239853
;   OTHER INFORMATION: exon N complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240569
;   OTHER INFORMATION: exon M1117 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240596
;   OTHER INFORMATION: exon M1090 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240617
;   OTHER INFORMATION: exon M1069 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240644
;   OTHER INFORMATION: exon MS2 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240824
;   OTHER INFORMATION: exon M862 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240994
;   OTHER INFORMATION: exon M692 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..241685
;   OTHER INFORMATION: exon M1 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240800..240993
;   OTHER INFORMATION: exon MS1 complement g34872 gene
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 241686..243685
;   OTHER INFORMATION: 5'regulatory region g34872 gene
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 290652..292652
;   OTHER INFORMATION: 3'regulatory region g34665 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 292653..292841

  Query Match             5.3%;  Score 53.4;  DB 4;  Length 319608;
  Best Local Similarity   52.5%;  Pred. No. 0.00019;
  Matches  117;  Conservative    0;  Mismatches  106;  Indels    0;  Gaps    0;

Qy       178 CCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGG 237
                ||| | | |   ||  ||| | | ||       |       || ||  ||| ||  |

Db     61813 CCTTCCTTCCATCTTTCCTTCCTTCTTTCCTTCCTTCCTTCCTTCTTTCTTCCTTCTTCC 61754

Qy       238 GCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 297
                | ||| ||  | | ||| ||| ||  ||  |||    |  || ||  ||    | |||

Db     61753 CTTTCTCCCTCCCTTCCTTCCTTTCTTCCTTACCTCAATCTTTCCTTCCTTCTTTCCTTC 61694

Qy       298 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 357
                ||| ||  ||  | | ||  || ||| ||  ||  | | | |  || ||  ||  | | |||

Db     61693 CTTCCTTCCTTCCTTCATTCCTTCCTTCCTTCCTTCCATCTTTCCTTCCTTCCTTCCTTC 61634

Qy       358 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
                ||| ||  ||  | | ||| ||| ||  ||  | | |||  ||

Db     61633 CTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTTTTT 61591


RESULT 20
US-09-679-409-1/c
; Sequence 1, Application US/09679409
; Patent No. 6555316
; GENERAL INFORMATION:
;  APPLICANT:  Cohen, Daniel
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Chumakov, Ilya
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Essioux, Laurent
;  TITLE OF INVENTION:  SCHIZOPHRENIA ASSOCIATED GENE, PROTEINS AND BIALLELIC MARKERS
;  FILE REFERENCE:  53.US15.CIP
;  CURRENT APPLICATION NUMBER: US/09/679,409
;  CURRENT FILING DATE:  2000-10-03
;  PRIOR APPLICATION NUMBER:  09/539,333
;  PRIOR FILING DATE:  2000-03-03
;  PRIOR APPLICATION NUMBER:  09/416,384
;  PRIOR FILING DATE:  1999-10-12
;  PRIOR APPLICATION NUMBER:  60/168,088
;  PRIOR FILING DATE:  1999-11-30
;  NUMBER OF SEQ ID NOS:  134
;  SOFTWARE:  Patent.pm
; SEQ ID NO 1
;   LENGTH: 319608
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 199122..201122
;   OTHER INFORMATION: 5'regulatory region
;   NAME/KEY: exon
;   LOCATION: 201123..201234
;   OTHER INFORMATION: exon S
;   NAME/KEY: exon
;   LOCATION: 201123..201560
;   OTHER INFORMATION: exon S2
;   NAME/KEY: exon
;   LOCATION: 214676..214793
;   OTHER INFORMATION: exon T
;   NAME/KEY: exon
;   LOCATION: 215702..215746
;   OTHER INFORMATION: exon U
;   NAME/KEY: exon
;   LOCATION: 216836..216994
;   OTHER INFORMATION: exon V
;   NAME/KEY: exon
;   LOCATION: 216836..217077
;   OTHER INFORMATION: exon V2
;   NAME/KEY: exon
;   LOCATION: 217671..217764
;   OTHER INFORMATION: exon V1
;   NAME/KEY: exon
;   LOCATION: 227655..227736
;   OTHER INFORMATION: exon V4
;   NAME/KEY: exon
;   LOCATION: 238715..238919
;   OTHER INFORMATION: exon V3
;   NAME/KEY: exon
```

```
;   LOCATION: 240440..240673
;   OTHER INFORMATION: exon W
;   NAME/KEY: exon
;   LOCATION: 240440..241153
;   OTHER INFORMATION: exon W2
;   NAME/KEY: exon
;   LOCATION: 241072..241291
;   OTHER INFORMATION: exon X
;   NAME/KEY: exon
;   LOCATION: 244353..244561
;   OTHER INFORMATION: exon Y
;   NAME/KEY: exon
;   LOCATION: 246273..247802
;   OTHER INFORMATION: exon Z
;   NAME/KEY: misc_feature
;   LOCATION: 247803..249803
;   OTHER INFORMATION: 3'regulatory region
;   NAME/KEY: allele
;   LOCATION: 8316
;   OTHER INFORMATION: 99-27943-150  : polymorphic base G or C
;   NAME/KEY: allele
;   LOCATION: 21672
;   OTHER INFORMATION: 99-27935-193  : polymorphic base G or C
;   NAME/KEY: allele
;   LOCATION: 65485
;   OTHER INFORMATION: 8-128-33  : polymorphic base C or T
;   NAME/KEY: allele
;   LOCATION: 95396
;   OTHER INFORMATION: 99-31960-363  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 107281
;   OTHER INFORMATION: 99-24656-260  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 160640
;   OTHER INFORMATION: 99-24639-163  : polymorphic base A or C
;   NAME/KEY: allele
;   LOCATION: 160876
;   OTHER INFORMATION: 99-24634-108  : polymorphic base A or T
;   NAME/KEY: allele
;   LOCATION: 168974
;   OTHER INFORMATION: 99-7652-162  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 170810
;   OTHER INFORMATION: 99-16100-147  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 173358
;   OTHER INFORMATION: 99-5862-167  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 189957
;   OTHER INFORMATION: 99-5919-215  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 197163
;   OTHER INFORMATION: 99-24658-410  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 200778
;   OTHER INFORMATION: 8-303-235  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 202651
;   OTHER INFORMATION: 8-300-221  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 202679
;   OTHER INFORMATION: 8-300-193  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 203378
;   OTHER INFORMATION: 8-299-128  : polymorphic base A or T
;   NAME/KEY: allele
;   LOCATION: 204138
;   OTHER INFORMATION: 8-296-213  : polymorphic base A or T
;   NAME/KEY: allele
;   LOCATION: 204605
;   OTHER INFORMATION: 8-252-190  : polymorphic base C or T
;   NAME/KEY: allele
;   LOCATION: 204934
;   OTHER INFORMATION: 99-24644-194  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 205206
;   OTHER INFORMATION: 8-295-248  : polymorphic base A or C
;   NAME/KEY: allele
;   LOCATION: 205329
;   OTHER INFORMATION: 8-295-125  : polymorphic base C or T
;   NAME/KEY: allele
;   LOCATION: 206064
;   OTHER INFORMATION: 8-293-130  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 206545
;   OTHER INFORMATION: 8-292-198  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 207313
;   OTHER INFORMATION: 8-251-322  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 208285
;   OTHER INFORMATION: 8-289-322  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 208960
;   OTHER INFORMATION: 8-287-249  : polymorphic base C or T
;   NAME/KEY: allele
;   LOCATION: 209123
;   OTHER INFORMATION: 8-287-86  : polymorphic base A or T
;   NAME/KEY: allele
;   LOCATION: 209631
;   OTHER INFORMATION: 8-285-319  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 210361
;   OTHER INFORMATION: 8-283-278  : polymorphic base G or C
;   NAME/KEY: allele
;   LOCATION: 210463
;   OTHER INFORMATION: 8-283-176  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 210486
;   OTHER INFORMATION: 8-283-153  : polymorphic base G or C
;   NAME/KEY: allele
;   LOCATION: 210583
;   OTHER INFORMATION: 8-283-56  : polymorphic base C or T
;   NAME/KEY: allele
;   LOCATION: 210879
;   OTHER INFORMATION: 8-282-345  : polymorphic base G or C
;   NAME/KEY: allele
;   LOCATION: 210964
;   OTHER INFORMATION: 8-282-260  : polymorphic base G or T
;   NAME/KEY: allele
;   LOCATION: 210979
;   OTHER INFORMATION: 8-282-245  : polymorphic base A or C
;   NAME/KEY: allele
;   LOCATION: 211050
;   OTHER INFORMATION: 8-282-174  : variable motif AAAGG or GAAGGAAGGAAGGAAGGAAGA
;   NAME/KEY: allele
;   LOCATION: 211132
;   OTHER INFORMATION: 8-282-92  : polymorphic base A or T
;   NAME/KEY: allele
;   LOCATION: 211247
;   OTHER INFORMATION: 8-281-367  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 211315
;   OTHER INFORMATION: 8-281-299  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 211366
;   OTHER INFORMATION: 8-281-248  : polymorphic base G or C
;   NAME/KEY: allele
;   LOCATION: 212520
;   OTHER INFORMATION: 8-279-197  : polymorphic base A or C
;   NAME/KEY: allele
;   LOCATION: 212821
;   OTHER INFORMATION: 8-278-289  : polymorphic base C or T
;   NAME/KEY: allele

Query Match             5.3%;  Score 53.4;  DB 4;  Length 319608;
```

```
  Best Local Similarity   52.5%;  Pred. No. 0.00019;
  Matches  117;  Conservative    0;  Mismatches  106;  Indels    0;  Gaps    0;

Qy        178 CCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGG 237
              ||| | | |   ||  ||| | | ||        |       || ||  ||| ||  |
Db      61818 CCTTCCTTCCATCTTTCCTTCCTTCTTTCCTTCCTTCCTTCTTTCTTCCTTCTTCC 61754

Qy        238 GCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 297
              | ||| ||  | | ||| ||| ||  ||  |||   |  || ||  ||    | |||
Db      61753 CTTTCTCCCTCCCTTCCTTCCTTTCTTCCTTACCTCAATCTTTCCTTCCTTCTTTCCTTC 61694

Qy        298 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 357
              ||| ||  ||  | |  || ||| ||  ||  | | | |  || ||  ||  | | |||
Db      61693 CTTCCTTCCTTCCTTCATTCCTTCCTTCCTTCCTTCCATCTTTCCTTCCTTCCTTCCTTC 61634

Qy        358 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
              ||| ||  ||  | | ||| ||| ||  ||  | | |||  ||
Db      61633 CTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTTTTT 61591


RESULT 21
US-08-595-937A-1/c
; Sequence 1, Application US/08595937A
; Patent No. 5919919
; Patent No. 5919919 5786467
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  A Brassica sp. Gene Promoter Highly Expressed
;    TITLE OF INVENTION:  During Tapetum Development
;    NUMBER OF SEQUENCES:  2
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/595,937A
;      FILING DATE:
;      CLASSIFICATION:  800
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1960 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-595-937A-1

  Query Match             5.3%;  Score 53;  DB 2;  Length 1960;
  Best Local Similarity   45.2%;  Pred. No. 3.8e-05;
  Matches  285;  Conservative    0;  Mismatches  335;  Indels   11;  Gaps    2;

Qy        116 TGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTA 175
              | ||  |  || ||  |  ||   |      |||| ||  || ||||||      |||
Db        835 TTCTCTCCAGCCGGTTTATTTTTCGGTTTTCCAGTCGGTTTGTTTGCTGCCTCTCCAGCC 776

Qy        176 GGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTA 235
              ||  |  ||||   |   || |  | | |   |   ||    ||| |||    |  | |
Db        775 GGTTTATTTGCAGCCCCTCCAGTCGGTTTATTTGTTGGTTTATCTGTTGGTCCTCCGGCA 716

Qy        236 GGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 295
              ||  |      |||  |   | ||    | |  ||  | ||   || | |    || ||
Db        715 GGTTTATCTTTTGGTTTATCTACTGATCCTCCAGCTGGTTTATCTGTTGGTCCACCAGCG 656

Qy        296 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 355
              | ||      |||  |  ||| |  ||| |  ||  | || ||| || |||   ||| |
Db        655 GGTTTATCTTTTGGTTTATTTGATGCCTCTCCAGCTGGTTTACTTTCTGGCTTATCTGTT 596

Qy        356 TGCTTGCTCGCTGGCCT--------GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 407
              |    |  ||  | |        | ||   || |   |  || ||  | || ||  ||
Db        595 GGTCCACCAGCGAGTTTATCTTTTGGTTTATGTGATGTTTCTCCAGCCGGTTTACTTTCT 536

Qy        408 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCC 467
              ||| |   || ||      |||  ||  |   |  | | | | | || | | | || ||
Db        535 GGCTTATCTGTTGGTCCACCAGCGGGTTTATCTTTTGGTTTATGTGATGTTTCTCCAGCT 476

Qy        468 TGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCT---GCTGGGCTGCTAGTAG 524
              | |  |||  | |      || ||| |  | |   |    |     | || |  |   |
Db        475 GGTTTACTTTCTGGCTTATTTGTTGGTCCTCCAACGGATTTATCTTTTGGTTTATTTGAT 416

Qy        525 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 584
              | ||      ||| || | |   ||| ||  |  | ||  ||  |         | |  |
Db        415 GTCTCTCCAGCTTCTGGCATATCTGTTGATCCACCAGCGGGTTTATCTTTTGGTTTATTT 356

Qy        585 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 644
              |    ||   |  | |||     | |||   |    || || |  |  |   | |  |
Db        355 GATGCCTCTCCAACCAGTTTACTTTCTGTCTTATCTGTTGGTCCTCCAGCGGGTTTATCA 296

Qy        645 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTT 704
              | | |    | || |     ||| ||| |  | ||  ||  |  |  ||||||||| |||
Db        295 ACCGGTTTACTTCCTGGCTTATCTGTTGGTCCTCCAGCGGGTTTATCTGCTGGCTTACTT 236

Qy        705 GCTTGCTTGCTAGTGGGGCCGCTTGCCTGCT 735
              || | ||    || || || |  ||  | |
Db        235 TCTGGTTTATCCGTTGGTCCTCCAGCGGGTT 205


RESULT 22
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14
```

```
  Query Match             5.3%;  Score 52.6;  DB 1;  Length 7218;
  Best Local Similarity   8.0%;  Pred. No. 7.9e-05;
  Matches   19;  Conservative  137;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        183 TTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGC 242
              || |  ||| | ||||  |  |||||  |    |  :: ::  ::: ::  :    :: :
Db       1031 TTCCGAGCTTGGCTGCAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYY 1090

Qy        243 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 302
              :: ::  ::: ::: ::: ::: ::  ::: ::: ::: ::: ::  ::: ::: ::: :
Db       1091 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1150

Qy        303 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 362
              :: ::  ::: ::: ::: ::: ::  ::: ::: ::: ::: ::  ::: ::: ::: :
Db       1151 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1210

Qy        363 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              :: ::  ::: ::: ::: ::: ::  ::: ::: ::: ::: ::  ::: ::: ::
Db       1211 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1267


RESULT 23
US-09-007-005-17/c
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

  Query Match             5.2%;  Score 51.8;  DB 3;  Length 289;
  Best Local Similarity   5.4%;  Pred. No. 4e-05;
  Matches   11;  Conservative  103;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        219 CTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 278
              |:|:   : : :| |:     ::: : :   : ::: : : : : :::   ::: : :::
Db        234 CYGYCYAYGYCYTYGYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSY 175

Qy        279 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 338
              : : : :  :: : : : : ::: : :   : ::: : : : : :::   ::: : :::
Db        174 NYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSY 115

Qy        339 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 398
              : : : :  :: : : : : ::: : :   : ::: : : : : :::   ::: : : :
Db        114 NYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYCY 55

Qy        399 TTGCTCGCTGGCCTGCTTGCTGA 421
              : :|:|:|  : : :|:|:| |
Db         54 AYTYTYGYTYAYAYTYTYGYTYA 32


RESULT 24
US-09-244-796-17/c
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350007
;  CURRENT APPLICATION NUMBER: US/09/244,796
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  EARLIER APPLICATION NUMBER: 09/007,005
;  EARLIER FILING DATE: 1998-01-14
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

  Query Match             5.2%;  Score 51.8;  DB 3;  Length 289;
  Best Local Similarity   5.4%;  Pred. No. 4e-05;
  Matches   11;  Conservative  103;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        219 CTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 278
              |:|:   : : :| |:     ::: : :   : ::: : : : : :::   ::: : :::
Db        234 CYGYCYAYGYCYTYGYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSY 175

Qy        279 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 338
              : : : :  :: : : : : ::: : :   : ::: : : : : :::   ::: : :::
Db        174 NYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSY 115

Qy        339 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 398
              : : : :  :: : : : : ::: : :   : ::: : : : : :::   ::: : : :
Db        114 NYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYCY 55

Qy        399 TTGCTCGCTGGCCTGCTTGCTGA 421
              : :|:|:|  : : :|:|:| |
Db         54 AYTYTYGYTYAYAYTYTYGYTYA 32


RESULT 25
US-09-019-095A-7/c
; Sequence 7, Application US/09019095A
; Patent No. 6287858
; GENERAL INFORMATION:
;  APPLICANT:  D'Andrea, Alan D.
;  APPLICANT:  Zhu, Yuan
;  TITLE OF INVENTION:  Deubiquitinating Enzymes That Regulate
;  TITLE OF INVENTION:  Cell Growth
;  FILE REFERENCE:  DFCI-435p2A2
;  CURRENT APPLICATION NUMBER: US/09/019,095A
;  CURRENT FILING DATE:  1998-02-05
;  PRIOR APPLICATION NUMBER:  PCT/US96/12884
;  PRIOR FILING DATE:  1996-08-07
;  PRIOR APPLICATION NUMBER:  US 60/002,066
;  PRIOR FILING DATE:  1995-08-09
;  PRIOR APPLICATION NUMBER:  US 60/019,787
;  PRIOR FILING DATE:  1996-06-14
;  NUMBER OF SEQ ID NOS:  51
```

```
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 1559
;   TYPE: DNA
;   ORGANISM: Murine
US-09-019-095A-7

  Query Match             5.1%;  Score 50.8;  DB 3;  Length 1559;
  Best Local Similarity   51.8%;  Pred. No. 0.00014;
  Matches  115;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy        198 CTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTG 257
              |||||| |    |||| |    |      |   ||||    | |  ||  |  | |
Db       1195 CTTGCTAACCACACAGAACCAGTATAAACAGCCTAGGCCTGGTTTTCCTTCCCTTTTGCC 1136

Qy        258 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 317
              ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | |||
Db       1135 CTTTCTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 1076

Qy        318 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 377
              ||| ||   |  |   ||  ||   || ||  |   ||  ||  ||| ||  | | |||
Db       1075 CTTTCTTTATCTCTCTCTCTCTTTCTCTTTCTCTCTCTCTATCTCTCTCTCTTCCTTC 1016

Qy        378 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              ||| |   |   | | ||| ||| ||  ||  | | ||| ||
Db       1015 CTTCCGTCCGTCCTTCCTTCCTTCCTTCCTTCCTTTCTTTCT 974


RESULT 26
US-09-318-443-7/c
; Sequence 7, Application US/09318443
; Patent No. 6197947
; GENERAL INFORMATION:
;  APPLICANT: Hemmati-Brivanlou, Ali
;  APPLICANT: Weinstein, Daniel C.
;  TITLE OF INVENTION: TRANSLATION INITIATION FACTOR 4AIII, AND METHODS OF USE
;  TITLE OF INVENTION: THEREOF
;  FILE REFERENCE: 600-1-211 N
;  CURRENT APPLICATION NUMBER: US/09/318,443
;  CURRENT FILING DATE: 1999-05-25
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 1682
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-318-443-7

  Query Match             5.0%;  Score 50.4;  DB 3;  Length 1682;
  Best Local Similarity   59.0%;  Pred. No. 0.00018;
  Matches  105;  Conservative    0;  Mismatches   71;  Indels    2;  Gaps    1;

Qy        238 GCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 297
              || |  | |    | |||  |  |||||||  | |||  |  |||||||  | |||
Db        177 GCCGACCTCGCTGCCGCTGCCGACCTCGCTGTGCCGCTGCCGACCTCGCTG--CCGCTGC 120

Qy        298 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 357
              |   |||||||  | |||  |  |||||||  | |||  |  |||||||  | |||
Db        119 CGACCTCGCTGTGCCGCTGCCGACCTCGCTGTGCCGCTGCCGACCTCGCTGTGCCGCTGC 60

Qy        358 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 415
              |   |||||||  | |||  |  |||||||  | |||  |  |||||||  | |||
Db         59 CGACCTCGCTGTGCCGCTGCCGACCTCGCTGTGCCGCTGCCGACCTCGCTGTGCCGCT 2


RESULT 27
US-08-623-906A-14/c
; Sequence 14, Application US/08623906A
; Patent No. 5874217
;  GENERAL INFORMATION:
;    APPLICANT:  Stevenson, Tamara
;    APPLICANT:  Dvorak, Jan
;    APPLICANT:  Halverson, Joy
;    TITLE OF INVENTION:  Microsatellite Sequences for Canine
;    TITLE OF INVENTION:  Genotyping
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/623,906A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sherwood, Pamela J.
;      REGISTRATION NUMBER:  36,677
;      REFERENCE/DOCKET NUMBER:  A-62282/BIR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  348 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..21
;      OTHER INFORMATION:  /note= "Nucleotides 1-21 are unique
;      OTHER INFORMATION:  flanking sequence"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  22..156
;      OTHER INFORMATION:  /note= "Nucleotides 22-156 are
;      OTHER INFORMATION:  repeat sequence"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  157..348
;      OTHER INFORMATION:  /note= "Nucleotides 157-348 are
;      OTHER INFORMATION:  unique flanking sequence"
US-08-623-906A-14

  Query Match             5.0%;  Score 49.8;  DB 2;  Length 348;
  Best Local Similarity   57.3%;  Pred. No. 0.00015;
  Matches   90;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        270 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 329
              | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||
Db        157 CTTTCTTTCTTCCTTTCTTTCTTTCTTCCTTTCTTCCTTCCTTCCTTCCTTCCTTT 98

Qy        330 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 389
              | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||   |
Db         97 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC 38

Qy        390 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 426
              |  || ||| ||  | || |   || |  || | |
Db         37 TTTCTTTCCTTTCTTTCAGGTCATTTTACAAAGAGCA 1


RESULT 28
US-09-759-359A-3
; Sequence 3, Application US/09759359A
; Patent No. 6492153
; GENERAL INFORMATION:
;  APPLICANT: ABU-THREIDEH, Jane et al
```

```
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001043
;  CURRENT APPLICATION NUMBER: US/09/759,359A
;  CURRENT FILING DATE:  2001-01-16
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 90541
;   TYPE: DNA
;   ORGANISM: Human
US-09-759-359A-3

  Query Match             5.0%;  Score 49.6;  DB 4;  Length 90541;
  Best Local Similarity   53.0%;  Pred. No. 0.0013;
  Matches  106;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy        220 TGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 279
              | ||  ||| ||         | ||| ||  | | ||| ||| |||  |        |   |
Db      34748 TCCTTCCTTCCTTCCTTCCTTCCTCTCTTTCTTTCTTTCTTTCTCTTTCTTTCTTTCTTT 34807

Qy        280 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 339
                ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |||  ||  |
Db      34808 CTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCCTTTTTCTT 34867

Qy        340 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 399
              |         | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||
Db      34868 TTTCTTTCTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 34927

Qy        400 TGCTCGCTGGCCTGCTTGCT 419
              | ||  ||  | | ||| ||
Db      34928 TTCTTTCTTTCTTCCTTTCT 34947


RESULT 29
US-09-268-992-7
; Sequence 7, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
;  APPLICANT: Chen, H.
;  APPLICANT: Freimer, N.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
;  TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
;  FILE REFERENCE: 7853-138
;  CURRENT APPLICATION NUMBER: US/09/268,992
;  CURRENT FILING DATE: 1999-03-16
;  EARLIER APPLICATION NUMBER: 09/236,134
;  EARLIER FILING DATE: 1999-01-22
;  EARLIER APPLICATION NUMBER: 60/106,056
;  EARLIER FILING DATE: 1998-10-28
;  EARLIER APPLICATION NUMBER: 60/088,312
;  EARLIER FILING DATE: 1998-06-05
;  EARLIER APPLICATION NUMBER: 60/078,044
;  EARLIER FILING DATE: 1998-03-16
;  NUMBER OF SEQ ID NOS: 84
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 72604
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: all n positions
;   OTHER INFORMATION: n=a, c, g, or t
US-09-268-992-7

  Query Match             4.9%;  Score 49.2;  DB 4;  Length 72604;
  Best Local Similarity   60.5%;  Pred. No. 0.0015;
  Matches   98;  Conservative    0;  Mismatches   63;  Indels    1;  Gaps    1;

Qy        251 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 310
              || ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db       7174 CTCCTTCCTTCCTTCCTTCCTTCCTTTCTTTCTTTCT-CCTTCCTTCCTTCCTTCCTTCC 7232

Qy        311 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 370
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db       7233 TTCCTTCCTTCCTTCCTTCCTTCCTTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC 7292

Qy        371 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 412
              | ||| ||| ||  ||  ||| ||| ||| |   | |  ||
Db       7293 TTTCTTTCTTTCTTTCTTTCCTTCTTTCTTTCGACCAGTTCT 7334


RESULT 30
US-09-657-474-7
; Sequence 7, Application US/09657474
; Patent No. 6399762
; GENERAL INFORMATION:
;  APPLICANT: Chen, H.
;  APPLICANT:  Freimer, N.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
;  TITLE OF INVENTION:  AND TREATING CHROMOSOME-18p RELATED DISORDERS
;  FILE REFERENCE: 7853-138
;  CURRENT APPLICATION NUMBER: US/09/657,474
;  CURRENT FILING DATE:  2000-09-07
;  PRIOR APPLICATION NUMBER: 09/268,992
;  PRIOR FILING DATE: 1999-03-16
;  PRIOR APPLICATION NUMBER: 09/236,134
;  PRIOR FILING DATE: 1999-01-22
;  PRIOR APPLICATION NUMBER: 60/106,056
;  PRIOR FILING DATE: 1998-10-28
;  PRIOR APPLICATION NUMBER: 60/088,312
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/078,044
;  PRIOR FILING DATE: 1998-03-16
;  NUMBER OF SEQ ID NOS: 84
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 72604
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: all n positions
;   OTHER INFORMATION: n=a, c, g, or t
US-09-657-474-7

  Query Match             4.9%;  Score 49.2;  DB 4;  Length 72604;
  Best Local Similarity   60.5%;  Pred. No. 0.0015;
  Matches   98;  Conservative    0;  Mismatches   63;  Indels    1;  Gaps    1;

Qy        251 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 310
              || ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db       7174 CTCCTTCCTTCCTTCCTTCCTTCCTTTCTTTCTTTCT-CCTTCCTTCCTTCCTTCCTTCC 7232

Qy        311 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 370
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db       7233 TTCCTTCCTTCCTTCCTTCCTTCCTTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC 7292

Qy        371 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 412
              | ||| ||| ||  ||  ||| ||| ||| |   | |  ||
Db       7293 TTTCTTTCTTTCTTTCTTTCCTTCTTTCTTTCGACCAGTTCT 7334


RESULT 31
US-08-229-515A-13/c
; Sequence 13, Application US/08229515A
; Patent No. 5518885
;  GENERAL INFORMATION:
;    APPLICANT:  RAZIUDDIN
;    APPLICANT:  SARKAR, FAZLUL H
;    TITLE OF INVENTION:  ERBB2 PROMOTER BINDING PROTEIN IN
;    TITLE OF INVENTION:  NEOPLASTIC DISEASE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
```

```
;    ADDRESSEE:  NEEDLE & ROSENBERG PC
;      STREET:  127 Peachtree Street, Suite 1200
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  usa
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/229,515A
;      FILING DATE:  19 APR 1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PERRYMAN, DAVID G
;      REGISTRATION NUMBER:  33,438
;      REFERENCE/DOCKET NUMBER:  1414.608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  404-688-0770
;      TELEFAX:  404-688-9880
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  650 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-229-515A-13

  Query Match             4.9%;  Score 48.8;  DB 1;  Length 650;
  Best Local Similarity   73.8%;  Pred. No. 0.00035;
  Matches   62;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy        249 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 308
              ||||||    |  | ||  | |  ||||||||||| ||| || ||||||||||| |||
Db        183 GCCTGCCCTGTGCCCCGACGCCTAGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 124

Qy        309 GCCTGCTTGCTTGCTCGCTGGCCT 332
              || ||||||||||| || |  ||
Db        123 GCTTGCTTGCTTGCTTGCAGTTCT 100


RESULT 32
US-08-645-865-13/c
; Sequence 13, Application US/08645865
; Patent No. 5654406
;  GENERAL INFORMATION:
;    APPLICANT:  RAZIUDDIN
;    APPLICANT:  SARKAR, FAZLUL H
;    TITLE OF INVENTION:  ERBB2 PROMOTER BINDING PROTEIN IN
;    TITLE OF INVENTION:  NEOPLASTIC DISEASE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG PC
;      STREET:  127 Peachtree Street, Suite 1200
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  usa
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/645,865
;      FILING DATE:  14 MAY 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PERRYMAN, DAVID G
;      REGISTRATION NUMBER:  33,438
;      REFERENCE/DOCKET NUMBER:  1414.608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  404-688-0770
;      TELEFAX:  404-688-9880
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  650 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-645-865-13

  Query Match             4.9%;  Score 48.8;  DB 1;  Length 650;
  Best Local Similarity   73.8%;  Pred. No. 0.00035;
  Matches   62;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy        249 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 308
              ||||||    |  | ||  | |  ||||||||||| ||| || ||||||||||| |||
Db        183 GCCTGCCCTGTGCCCCGACGCCTAGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 124

Qy        309 GCCTGCTTGCTTGCTCGCTGGCCT 332
              || ||||||||||| || |  ||
Db        123 GCTTGCTTGCTTGCTTGCAGTTCT 100


RESULT 33
US-08-903-800A-3
; Sequence 3, Application US/08903800A
; Patent No. 5935789
;  GENERAL INFORMATION:
;    APPLICANT:  RHEE, Sang-Ki
;    APPLICANT:  CHOI, Eui-Sung
;    APPLICANT:  KIM, Chul-Ho
;    APPLICANT:  SOHN, Jung-Hoon
;    APPLICANT:  KANG, Hyun-Ah
;    APPLICANT:  KIM, Hwa-Young
;    TITLE OF INVENTION:  AUTONOMOUSLY REPLICATING SEQUENCES,
;    TITLE OF INVENTION:  GAPDH GENE AND PROMOTER DERIVED FROM
;    TITLE OF INVENTION:  HANSENULA POLYMORPHA, EXPRESSION
;    TITLE OF INVENTION:  VECTORS CONTAINING SAME AND METHOD FOR
;    TITLE OF INVENTION:  THE SELECTION OF TRANSFORMANTS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RHEE, Sang-Ki
;      STREET:  Keukdong Villa Ka-101, Kwangjang-dong,
;      STREET:  Kwangjin-gu
;      CITY:  Seoul
;      STATE:  Seoul
;      COUNTRY:  Republic of Korea
;      ZIP:  143-210
;      ADDRESSEE:  CHOI, Eui-Sung
;      STREET:  Dasol Apt. 102-507, #395-3, Kung-dong,
;      STREET:  Yuseong-gu
;      CITY:  Daejeon
;      STATE:  Daejeon
;      COUNTRY:  Republic of Korea
;      ZIP:  305-335
;      ADDRESSEE:  KIM, Chul-Ho
;      STREET:  Shindonga Apt. 2-1101, Yongjeon-dong, Dong-gu
;      CITY:  Daejeon
;      STATE:  Daejeon
;      COUNTRY:  Republic of Korea
;      ZIP:  300-200
;      ADDRESSEE:  SOHN, Jung-Hoon
;      STREET:  Nuri Apt. 103-506, Wolpyung-dong, Seo-gu
;      CITY:  Daejeon
;      STATE:  Daejeon
;      COUNTRY:  Republic of Korea
;      ZIP:  302-280
;      ADDRESSEE:  KANG, Hyun-Ah
;      STREET:  Kyungseong Keunmaul Apt. 125-1501, Kalma-dong,
;      STREET:  Seo-gu
;      CITY:  Daejeon
;      STATE:  Daejeon
```

```
;      COUNTRY:  Republic of Korea
;      ZIP:  302-171
;      ADDRESSEE:  KIM, Hwa-Young
;      STREET:  #3-462, Bukahyun-3-dong, Seodaemun-gu
;      CITY:  Seoul
;      STATE:  Seoul
;      COUNTRY:  Republic of Korea
;      ZIP:  120-190
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5inch, 1.44MB storage
;      COMPUTER:  IBM PC/AT
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/903,800A
;      FILING DATE:  31-JUL-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  KR 97-3173
;      FILING DATE:  31-JAN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:
;      REGISTRATION NUMBER:
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:
;      TELEFAX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  405 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      OTHER INFORMATION:  autonomously replicating sequnce of
;      OTHER INFORMATION:  Hansenula polymorpha DL-1(ATCC 26012)
US-08-903-800A-3

  Query Match             4.8%;  Score 48.4;  DB 2;  Length 405;
  Best Local Similarity   52.5%;  Pred. No. 0.00038;
  Matches  106;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        217 GGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 276
              ||  || | ||      | |||  ||  | ||||  | | ||  | | || ||   ||
Db        203 GGACGCGGTCTGAGGCCTGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGG 262

Qy        277 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 336
              | | ||  | ||||  | | ||  | | || ||   ||  | | ||  | ||||  | |
Db        263 GGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTG 322

Qy        337 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 396
              ||  | | || ||   ||  | | ||  | ||||  | | ||  | | || ||   ||
Db        323 GCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGG 382

Qy        397 GCTTGCTCGCTGGCCTGCTTGC 418
              | | ||  | ||||  | | ||
Db        383 GGTGGCGGGGTGGCGGGGTGGC 404


RESULT 34
US-09-252-991A-16386/c
; Sequence 16386, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER.OF SEQ ID NOS:  33142
; SEQ ID NO 16386
;   LENGTH: 951
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16386

  Query Match             4.8%;  Score 48.2;  DB 4;  Length 951;
  Best Local Similarity   51.6%;  Pred. No. 0.00058;
  Matches  213;  Conservative    0;  Mismatches  188;  Indels   12;  Gaps    4;

Qy         13 GCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCT 72
              ||   ||| |||   ||    |    | | | |   | ||  |   |  ||||  ||||
Db        881 GCAGCGGGCGCGCTCGAAGACGCGGCGGGAGCGGCCGGTTTCGCCGCGGCGGGCTTGGCT 822

Qy         73 GCCTGCCTGCTCGCTG--GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGC 130
              ||| || |  |||| |  | ||  ||       ||   |  |||  | ||   ||| |||
Db        821 GCCGGCTTCTTCGCCGCAGGCTTTTTCGCAGCCGGCTTGGCGGCAGGTTTCGCCGCGGGC 762

Qy        131 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGG---GACAGTAGGCCTGCTTGCT 187
               ||   || |  |   | |  ||| ||  ||| |  ||     ||| |||| |    ||
Db        761 TTGGCGGCGGTCGCAGCGGCCGGCTTGGCTGCGGCAGGTTTCGCAGCAGGCTTCGCCGCG 702

Qy        188 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCT 247
                || | ||||  |||  |  | | ||   ||||| | |||   ||| ||  ||   ||
Db        701 GCCTTGGCTGCCGGCTTGGTCGCCGGTTTCGCTGCGGCCTTCGCAGTCGGTTTGGCGGCG 642

Qy        248 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 307
              ||| ||  |||  ||| |  ||||    |  | ||        | | ||  ||   |||
Db        641 GGCTTGGCTGCCGGCTTGGCGGCCGCGGTTTTCGCCG------CGGGTTTTTCGCCGCG 588

Qy        308 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 367
              ||  |    ||  ||| | |||   ||||  |||  | |  ||||  ||||  ||||
Db        587 GGTTTCGCCGCAGGCTTGGCGGCGGCCTTGGCTGCCGGTTTAGCTGCCGGCTTGGCCGCT 528

Qy        368 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 420
              |   | | ||  ||  ||| | | ||  | ||   ||| ||| ||   ||||
Db        527 GCGGTTTTCGC-CGCGGGCTTGGCAGCCGGTTTCGCCGCAGGCTTGGCCGCTG 476


RESULT 35
US-09-252-991A-16212
; Sequence 16212, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 16212
;   LENGTH: 1092
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16212

  Query Match             4.8%;  Score 48.2;  DB 4;  Length 1092;
  Best Local Similarity   51.6%;  Pred. No. 0.00061;
  Matches  213;  Conservative    0;  Mismatches  188;  Indels   12;  Gaps    4;

Qy         13 GCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCT 72
              ||   ||| |||   ||    |    | | | |   | ||  |   |  ||||  ||||
Db        194 GCAGCGGGCGCGCTCGAAGACGCGGCGGGAGCGGCCGGTTTCGCCGCGGCGGGCTTGGCT 253

Qy         73 GCCTGCCTGCTCGCTG--GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGC 130
```

```
Db        254 GCCGGCTTCTTCGCCGCAGGCTTTTTCGCAGCCGGCTTGGCGGCAGGTTTCGCCGCGGGC 313

Qy        131 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGG---GACAGTAGGCCTGCTTGCT 187
Db        314 TTGGCGGCGGTCGCAGCGGCCGGCTTGGCTGCGGCAGGTTTCGCAGCAGGCTTCGCCGCG 373

Qy        188 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCT 247
Db        374 GCCTTGGCTGCCGGCTTGGTCGCCGGTTTCGCTGCGGCCTTCGCAGTCGGTTTGGCGGCG 433

Qy        248 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 307
Db        434 GGCTTGGCTGCCGGCTTGGCGGCCGCGGTTTCGCCG------CGGGTTTTTTCGCCGCG 487

Qy        308 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 367
Db        488 GGTTTCGCCGCAGGCTTGGCGGCGGCCTTGGCTGCCGGTTTAGCTGCCGGCTTGGCCGCT 547

Qy        368 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 420
Db        548 GCGGTTTTCGC-CGCGGGCTTGGCAGCCGGTTTCGCCGCAGGCTTGGCCGCTG 599


RESULT 36
US-09-252-991A-16493/c
; Sequence 16493, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 16493
;   LENGTH: 1311
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16493

  Query Match             4.8%;  Score 48.2;  DB 4;  Length 1311;
  Best Local Similarity   51.6%;  Pred. No. 0.00065;
  Matches  213;  Conservative    0;  Mismatches  188;  Indels   12;  Gaps    4;

Qy         13 GCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCT 72
Db        457 GCAGCGGGCGCGCTCGAAGACGCGGCGGGAGCGGCCGGTTTCGCCGCGGCGGGCTTGGCT 398

Qy         73 GCCTGCCTGCTCGCTG--GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGC 130
Db        397 GCCGGCTTCTTCGCCGCAGGCTTTTTCGCAGCCGGCTTGGCGGCAGGTTTCGCCGCGGGC 338

Qy        131 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGG---GACAGTAGGCCTGCTTGCT 187
Db        337 TTGGCGGCGGTCGCAGCGGCCGGCTTGGCTGCGGCAGGTTTCGCAGCAGGCTTCGCCGCG 278

Qy        188 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCT 247
Db        277 GCCTTGGCTGCCGGCTTGGTCGCCGGTTTCGCTGCGGCCTTCGCAGTCGGTTTGGCGGCG 218

Qy        248 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 307
Db        217 GGCTTGGCTGCCGGCTTGGCGGCCGCGGTTTTCGCCG------CGGGTTTTTTCGCCGCG 164

Qy        308 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 367
Db        163 GGTTTCGCCGCAGGCTTGGCGGCGGCCTTGGCTGCCGGTTTAGCTGCCGGCTTGGCCGCT 104

Qy        368 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 420
Db .      103 GCGGTTTTCGC-CGCGGGCTTGGCAGCCGGTTTCGCCGCAGGCTTGGCCGCTG 52


RESULT 37
US-09-252-991A-16005
; Sequence 16005, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 16005
;   LENGTH: 1530
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16005

  Query Match             4.8%;  Score 48.2;  DB 4;  Length 1530;
  Best Local Similarity   51.6%;  Pred. No. 0.00069;
  Matches  213;  Conservative    0;  Mismatches  188;  Indels   12;  Gaps    4;

Qy         13 GCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCT 72
Db        787 GCAGCGGGCGCGCTCGAAGACGCGGCGGGAGCGGCCGGTTTCGCCGCGGCGGGCTTGGCT 846

Qy         73 GCCTGCCTGCTCGCTG--GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGC 130
Db        847 GCCGGCTTCTTCGCCGCAGGCTTTTTCGCAGCCGGCTTGGCGGCAGGTTTCGCCGCGGGC 906

Qy        131 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGG---GACAGTAGGCCTGCTTGCT 187
Db        907 TTGGCGGCGGTCGCAGCGGCCGGCTTGGCTGCGGCAGGTTTCGCAGCAGGCTTCGCCGCG 966

Qy        188 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCT 247
Db        967 GCCTTGGCTGCCGGCTTGGTCGCCGGTTTCGCTGCGGCCTTCGCAGTCGGTTTGGCGGCG 1026

Qy        248 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 307
Db       1027 GGCTTGGCTGCCGGCTTGGCGGCCGCGGTTTTCGCCG------CGGGTTTTTTCGCCGCG 1080

Qy        308 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 367
Db       1081 GGTTTCGCCGCAGGCTTGGCGGCGGCCTTGGCTGCCGGTTTAGCTGCCGGCTTGGCCGCT 1140

Qy        368 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 420
Db       1141 GCGGTTTTCGC-CGCGGGCTTGGCAGCCGGTTTCGCCGCAGGCTTGGCCGCTG 1192


RESULT 38
US-09-252-991A-6106
; Sequence 6106, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
```

```
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 6106
;    LENGTH: 756
;    TYPE: DNA
;    ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6106

  Query Match             4.8%;  Score 48;  DB 4;  Length 756;
  Best Local Similarity   49.5%;  Pred. No. 0.00061;
  Matches  292;  Conservative    0;  Mismatches  285;  Indels   13;  Gaps    6;

Qy        143 GGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGC 202
              ||| | |       |||| | | ||     || | |     || |    | | |||  |
Db         13 GGGGCGGCGCCGTTGCTGCCCGCGGCAGGCGTTGCCGAAGCAGCGCTGGAGGCGGCTGCC 72

Qy        203 TGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGC--TT 260
              | | |     |    || || |||||   || |||   | | || ||| || ||||
Db         73 GGGGCGCTGGGAGTCGCCGCAGGCTTCGCAGCAGGAGCGGTGGCGGGCTTGGTTGCGGCA 132

Qy        261 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 320
              | | |  | || |  | ||  ||| ||| ||   ||  | |     ||  ||||
Db        133 GGTTTCGCGGCTACGGGTTTCGCCGCAGGCTTGGCCGCCGGTTTCGCCGCAGGCTTCGCA 192

Qy        321 GCTCGCTGGCCTGCTTGCTT---GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 377
              ||  ||| | ||||  ||||   |  ||| |   |     |  ||| |   ||| | ||
Db        193 GCCGGCTTGGCTGCAGGCTTCGCGGCCGCGGATTTCGCCACAGGCTTGGCCGCCGGTTTC 252

Qy        378 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTG 437
                ||  |||   | ||| ||||   || ||| |    |||| ||     |  | |  |
Db        253 GCCGCAGGCTTCGCAGCTGGCTTGGCTGCGGGCTTCGCGGCTGCGGTTTTCGCCGCGG-G 311

Qy        438 CTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTG 497
              ||||   |          ||  | || |  |||  ||||||   |  ||||| |  |||||
Db        312 CTTGGCCGTCGGTTTCGCGACAGGCTTCGCTGCCGGCTTGGCGG-CGGGCTTCGCGGCTG 370

Qy        498 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTG 557
              |   |   || || ||  ||   |  ||  |         |||     | | ||| |||||
Db        371 CGGTTTTCGCCGCAGGCTTGGCCGCCGGTTTCGCCACAGGCTTCGCCGCCGGCTTGGCTG 430

Qy        558 CTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTG 617
              |     ||||| |  ||||||   |   || || ||  ||   |  ||  ||   |   |
Db        431 C-----GGGCTTCGCGGCTGCGGTTTTCGCCGCGGGCTTGGCCGCCGGTTTCGCAGCAGG 485

Qy        618 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTG 677
              ||     | | ||| ||||||  |    ||||| | | |  | |   |   ||| ||| |
Db        486 CTTCGCCGTCGGCTTGGCTGCGGGCTTCGCTGCCGCGGTTTTCG-CCGCAGGCTTGGCCG 544

Qy        678 CTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 727
              |  ||   || ||  |||    ||  |||||   ||  |    ||||||
Db        545 CCGGTTTCGCAGCAGGCTTCGCAGCCGGCTTGGCCGCAGGCTTCGCCGCT 594


RESULT 39
US-09-252-991A-6186
; Sequence 6186, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 6186
;    LENGTH: 912
;    TYPE: DNA
;    ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6186

  Query Match             4.8%;  Score 48;  DB 4;  Length 912;
  Best Local Similarity   49.5%;  Pred. No. 0.00065;
  Matches  292;  Conservative    0;  Mismatches  285;  Indels   13;  Gaps    6;

Qy        143 GGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGC 202
              ||| | |       |||| | | ||     || | |     || |    | | |||  |
Db        207 GGGGCGGCGCCGTTGCTGCCCGCGGCAGGCGTTGCCGAAGCAGCGCTGGAGGCGGCTGCC 266

Qy        203 TGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGC--TT 260
              | | |     |    || || |||||   || |||   | | || ||| || ||||
Db        267 GGGGCGCTGGGAGTCGCCGCAGGCTTCGCAGCAGGAGCGGTGGCGGGCTTGGTTGCGGCA 326

Qy        261 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 320
              | | |  | || |  | ||  ||| ||| ||   ||  | |     ||  ||||
Db        327 GGTTTCGCGGCTACGGGTTTCGCCGCAGGCTTGGCCGCCGGTTTCGCCGCAGGCTTCGCA 386

Qy        321 GCTCGCTGGCCTGCTTGCTT---GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 377
              ||  ||| | ||||  ||||   |  ||| |   |     |  ||| |   ||| | ||
Db        387 GCCGGCTTGGCTGCAGGCTTCGCGGCCGCGGATTTCGCCACAGGCTTGGCCGCCGGTTTC 446

Qy        378 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTG 437
                ||  |||   | ||| ||||   || ||| |    |||| ||     |  | |  |
Db        447 GCCGCAGGCTTCGCAGCTGGCTTGGCTGCGGGCTTCGCGGCTGCGGTTTTCGCCGCGG-G 505

Qy        438 CTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTG 497
              ||||   |          ||  | || |  |||  ||||||   |  ||||| |  |||||
Db        506 CTTGGCCGTCGGTTTCGCGACAGGCTTCGCTGCCGGCTTGGCGG-CGGGCTTCGCGGCTG 564

Qy        498 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTG 557
              |   |   || || ||  ||   |  ||  |         |||     | | ||| |||||
Db        565 CGGTTTTCGCCGCAGGCTTGGCCGCCGGTTTCGCCACAGGCTTCGCCGCCGGCTTGGCTG 624

Qy        558 CTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTG 617
              |     ||||| |  ||||||   |   || || ||  ||   |  ||  ||   |   |
Db        625 C-----GGGCTTCGCGGCTGCGGTTTTCGCCGCGGGCTTGGCCGCCGGTTTCGCAGCAGG 679

Qy        618 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTG 677
              ||     | | ||| ||||||  |    ||||| | | |  | |   |   ||| ||| |
Db        680 CTTCGCCGTCGGCTTGGCTGCGGGCTTCGCTGCCGCGGTTTTCG-CCGCAGGCTTGGCCG 738

Qy        678 CTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTAGTGGGGCCGCT 727
              |  ||   || ||  |||    ||  |||||   ||  |    ||||||
Db        739 CCGGTTTCGCAGCAGGCTTCGCAGCCGGCTTGGCCGCAGGCTTCGCCGCT 788


RESULT 40
US-09-252-991A-6445/c
; Sequence 6445, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 6445
;    LENGTH: 915
;    TYPE: DNA
;    ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6445

  Query Match             4.8%;  Score 48;  DB 4;  Length 915;
  Best Local Similarity   49.5%;  Pred. No. 0.00065;
```

```
  Matches  292;  Conservative    0;  Mismatches  285;  Indels   13;  Gaps    6;

Qy        143 GGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGC 202
              ||| | |      ||||  | | ||     || | |      || |    | | ||| |
Db        720 GGGGCGGCGCCGTTGCTGCCCGCGGCAGGCGTTGCCGAAGCAGCGCTGGAGGCGGCTGCC 661

Qy        203 TGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGC--TT 260
               | | |    |    || || |||||   || |||   | | || ||| || ||||
Db        660 GGGGCGCTGGGAGTCGCCGCAGGCTTCGCAGCAGGAGCGGTGGCGGGCTTGGTTGCGGCA 601

Qy        261 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 320
              | |  |  | || |  | ||   ||| ||| ||   ||  | |     ||  ||||
Db        600 GGTTTCGCGGCTACGGGTTTCGCCGCAGGCTTGGCCGCCGGTTTCGCCGCAGGCTTCGCA 541

Qy        321 GCTCGCTGGCCTGCTTGCTT---GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 377
              ||  ||| | ||||  |||   |  ||| |   |     |  ||| |- ||| | ||
Db        540 GCCGGCTTGGCTGCAGGCTTCGCGGCCGCGGATTTCGCCACAGGCTTGGCCGCCGGTTTC 481

Qy        378 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTG 437
                 ||  |||   | ||| |||||   || ||| |    ||||  ||     |  | |  |
Db        480 GCCGCAGGCTTCGCAGCTGGCTTGGCTGCGGGCTTCGCGGCTGCGGTTTTCGCCGCGG-G 422

Qy        438 CTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTG 497
              ||||   |          ||  | || |  |||  ||||||   |  |||||  |  |||||
Db        421 CTTGGCCGTCGGTTTCGCGACAGGCTTCGCTGCCGGCTTGGCGG-CGGGCTTCGCGGCTG 363

Qy        498 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTG 557
              |   |  || || ||  ||   |  ||  |          |||     | | ||| |||||
Db        362 CGGTTTTCGCCGCAGGCTTGGCCGCCGGTTTCGCCACAGGCTTCGCCGCCGGCTTGGCTG 303

Qy        558 CTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTG 617
              |     |||||  |  ||||||    |   || || ||  ||    |  ||  ||   |  |
Db        302 C-----GGGCTTCGCGGCTGCGGTTTTCGCCGCGGGCTTGGCCGCCGGTTTCGCAGCAGG 248

Qy        618 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTG 677
              ||      | | |||  ||||||  |    |||||  | |  |  | |   |   ||| ||| |
Db        247 CTTCGCCGTCGGCTTGGCTGCGGGCTTCGCTGCCGCGGTTTTCG-CCGCAGGCTTGGCCG 189

Qy        678 CTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 727
              |  ||   || ||  |||     ||  |||||   ||  |    ||||||
Db        188 CCGGTTTCGCAGCAGGCTTCGCAGCCGGCTTGGCCGCAGGCTTCGCCGCT 139


RESULT 41
US-09-252-991A-6282/c
; Sequence 6282, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 6282
;   LENGTH: 1200
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6282

  Query Match             4.8%;  Score 48;  DB 4;  Length 1200;
  Best Local Similarity   49.5%;  Pred. No. 0.00072;
  Matches  292;  Conservative    0;  Mismatches  285;  Indels   13;  Gaps    6;

Qy        143 GGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGC 202
              ||| | |      ||||  | | ||     || | |      || |    | | ||| |
Db       1184 GGGGCGGCGCCGTTGCTGCCCGCGGCAGGCGTTGCCGAAGCAGCGCTGGAGGCGGCTGCC 1125

Qy        203 TGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGC--TT 260
               | | |    |    || || |||||   || |||   | | || ||| || ||||
Db       1124 GGGGCGCTGGGAGTCGCCGCAGGCTTCGCAGCAGGAGCGGTGGCGGGCTTGGTTGCGGCA 1065

Qy        261 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 320
              | |  |  | || |  | ||   ||| ||| ||   ||  | |     ||  ||||
Db       1064 GGTTTCGCGGCTACGGGTTTCGCCGCAGGCTTGGCCGCCGGTTTCGCCGCAGGCTTCGCA 1005

Qy        321 GCTCGCTGGCCTGCTTGCTT---GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 377
              ||  ||| | ||||  |||   |  ||| |   |     |  ||| |  ||| | ||
Db       1004 GCCGGCTTGGCTGCAGGCTTCGCGGCCGCGGATTTCGCCACAGGCTTGGCCGCCGGTTTC 945

Qy        378 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTG 437
                 ||  |||   | ||| |||||   || ||| |    ||||  ||     |  | |  |
Db        944 GCCGCAGGCTTCGCAGCTGGCTTGGCTGCGGGCTTCGCGGCTGCGGTTTTCGCCGCGG-G 886

Qy        438 CTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTG 497
              ||||   |          ||  | || |  |||  ||||||   |  |||||  |  |||||
Db        885 CTTGGCCGTCGGTTTCGCGACAGGCTTCGCTGCCGGCTTGGCGG-CGGGCTTCGCGGCTG 827

Qy        498 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTG 557
              |   |  || || ||  ||   |  ||  |          |||     | | ||| |||||
Db        826 CGGTTTTCGCCGCAGGCTTGGCCGCCGGTTTCGCCACAGGCTTCGCCGCCGGCTTGGCTG 767

Qy        558 CTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTG 617
              |     |||||  |  ||||||    |   || || ||  ||    |  ||  ||   |  |
Db        766 C-----GGGCTTCGCGGCTGCGGTTTTCGCCGCGGGCTTGGCCGCCGGTTTCGCAGCAGG 712

Qy        618 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTG 677
              ||      | | |||  ||||||  |    |||||  | |  |  | |   |   ||| ||| |
Db        711 CTTCGCCGTCGGCTTGGCTGCGGGCTTCGCTGCCGCGGTTTTCG-CCGCAGGCTTGGCCG 653

Qy        678 CTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 727
              |  ||   || ||  |||     ||  |||||   ||  |    ||||||
Db        652 CCGGTTTCGCAGCAGGCTTCGCAGCCGGCTTGGCCGCAGGCTTCGCCGCT 603


RESULT 42
US-09-252-991A-6023
; Sequence 6023, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 6023
;   LENGTH: 1806
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6023

  Query Match             4.8%;  Score 48;  DB 4;  Length 1806;
  Best Local Similarity   49.5%;  Pred. No. 0.00083;
  Matches  292;  Conservative    0;  Mismatches  285;  Indels   13;  Gaps    6;

Qy        143 GGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGC 202
              ||| | |      ||||  | | ||     || | |      || |    | | ||| |
Db        551 GGGGCGGCGCCGTTGCTGCCCGCGGCAGGCGTTGCCGAAGCAGCGCTGGAGGCGGCTGCC 610

Qy        203 TGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGC--TT 260
               | | |    |    || || |||||   || |||   | | || ||| || ||||
Db        611 GGGGCGCTGGGAGTCGCCGCAGGCTTCGCAGCAGGAGCGGTGGCGGGCTTGGTTGCGGCA 670

Qy        261 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 320
              | |  |  | || |  | ||   ||| ||| ||   ||  | |     ||  ||||
```

```
Db        671 GGTTTCGCGGCTACGGGTTTCGCCGCAGGCTTGGCCGCCGGTTTCGCCGCAGGCTTCGCA 730

Qy        321 GCTCGCTGGCCTGCTTGCTT---GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 377
              ||  ||| | ||||  ||||   |  ||| |   |     |  ||| |   ||| | ||
Db        731 GCCGGCTTGGCTGCAGGCTTCGCGGCCGCGGATTTCGCCACAGGCTTGGCCGCCGGTTTC 790

Qy        378 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTG 437
                ||  |||   | ||| |||||   || ||| |    |||| ||      |  | |  |
Db        791 GCCGCAGGCTTCGCAGCTGGCTTGGCTGCGGGCTTCGCGGCTGCGGTTTTCGCCGCGG-G 849

Qy        438 CTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTG 497
              ||||   |          ||  | || |  |||  ||||||   |  ||||| |  |||||
Db        850 CTTGGCCGTCGGTTTCGCGACAGGCTTCGCTGCCGGCTTGGCGG-CGGGCTTCGCGGCTG 908

Qy        498 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTG 557
              |   |   || || ||  ||   |  ||  |         |||     | | ||| |||||
Db        909 CGGTTTTCGCCGCAGGCTTGGCCGCCGGTTTCGCCACAGGCTTCGCCGCCGGCTTGGCTG 968

Qy        558 CTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTG 617
              |     ||||| |  |||||   |   || || ||  ||    |  ||  ||   |   |
Db        969 C-----GGGCTTCGCGGCTGCGGTTTTCGCCGCGGGCTTGGCCGCCGGTTTCGCAGCAGG 1023

Qy        618 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTG 677
              ||     | | ||| ||||||  |    ||||| | | |  | |   |   ||| ||| |
Db       1024 CTTCGCCGTCGGCTTGGCTGCGGGCTTCGCTGCCGCGGTTTTCG-CCGCAGGCTTGGCCG 1082

Qy        678 CTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 727
              |  ||   || ||  |||     ||  |||||   ||  |    ||||||
Db       1083 CCGGTTTCGCAGCAGGCTTCGCAGCCGGCTTGGCCGCAGGCTTCGCCGCT 1132


RESULT 43
US-09-579-664B-5
; Sequence 5, Application US/09579664B
; Patent No. 6514719
; GENERAL INFORMATION:
;  APPLICANT: Immunex Corporation
;  APPLICANT:  Bird, Timothy A.
;  APPLICANT:  Virca, G. Duke
;  APPLICANT:  Martin, Unja
;  APPLICANT:  Anderson, Dirk M.
;  TITLE OF INVENTION: NOVEL MURINE AND HUMAN KINASES
;  FILE REFERENCE: 2923-A
;  CURRENT APPLICATION NUMBER: US/09/579,664B
;  CURRENT FILING DATE:  2000-05-26
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
;   LENGTH: 3228
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-579-664B-5

  Query Match             4.8%;  Score 47.8;  DB 4;  Length 3228;
  Best Local Similarity   60.3%;  Pred. No. 0.0012;
  Matches   79;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy        254 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 313
              ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | |
Db       3096 CTTTCTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 3155

Qy        314 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 373
              ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | |
Db       3156 CTTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 3215

Qy        374 CTTGCTTGCTC 384
              |||  ||  ||
Db       3216 CTTCTTTCTTC 3226


RESULT 44
US-09-078-294-6/c
; Sequence 6, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT: Du Sart, Desiree
;  APPLICANT: Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/078,294
;  CURRENT FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 18443
;   TYPE: DNA
;   ORGANISM: BAC-F2 contig 2
US-09-078-294-6

  Query Match             4.7%;  Score 47;  DB 3;  Length 18443;
  Best Local Similarity   50.2%;  Pred. No. 0.0036;
  Matches  116;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy        250 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 309
              ||| ||| | | |   |   ||| | | | | ||  ||  | |  || ||  ||  |||
Db      10872 CCTTCTTTCCTTCCTCCCTCCCTTCCTTCCTTCTTCCTTCCTTCTTTTCTGTCTTCCTGC 10813

Qy        310 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 369
              |||  |  ||| |   ||  | | ||| ||| ||  |   | | | | ||| ||  ||
Db      10812 CCTCTTGCCTTCCCTCCTTCCTTCCTTCCTTTCTTCCCTCCTTTCCTTCTTCCTTTCTTC 10753

Qy        370 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 429
              ||| ||| | | ||  ||   || ||| | | ||| |   |||     ||   |  ||
Db      10752 CCTCCTTTCCTTCTTCCTTTTCTTCTTTCCTTCTCTCCTTCCTTTCCTCTATAGATACTT 10693

Qy        430 TAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 480
                |  |   ||     ||  | ||    |  ||||| |   ||   | ||
Db      10692 AGTGATTCTCTGTGATATACTGAGATAAGATTCCTGCTTTTGGGAAGGTTA 10642


RESULT 45
US-08-332-766A-26/c
; Sequence 26, Application US/08332766A
; Patent No. 5843647
;  GENERAL INFORMATION:
;    APPLICANT:  JEFFREYS, Alec J.
;    APPLICANT:  ARMOUR, John
;    TITLE OF INVENTION:  SIMPLE TANDEM REPEATS
;    NUMBER OF SEQUENCES:  125
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D. C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/332,766A
;      FILING DATE:  01-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9326052.9
;      FILING DATE:  21-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BIRD, Donald J.
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  217211/M94/0434/GB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
```

```
;    TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  446 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-332-766A-26

  Query Match             4.7%;  Score 46.8;  DB 2;  Length 446;
  Best Local Similarity   54.7%;  Pred. No. 0.0011;
  Matches   93;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy        250 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 309
              ||| ||| ||| |   |   |||   | |   ||  |   ||| | |.| | ||  |
Db        354 CCTTCTTTCTTTCCTTCCTTCCTTTCTTCCCTCTTTCCTTCCTTCCTTCCTTCTTTCCTT 295

Qy        310 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 369
              ||| | | | | | | |   ||| ||| ||| ||  |   | |  || ||| ||  ||
Db        294 CCTTCCTTCCTTCCCTCCTCCCTCCTTCCTTCCTTCCCTCCTTCTTTCCTTCCTTCCTTC 235

Qy        370 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              ||| |||  || ||  ||  | | ||  ||| ||| ||  | | ||| ||
Db        234 CCTCCTTCTTTCCTTCCTTCCTTCCTCTCTTCCTCCCTCCCTTCCTTCCT 185


RESULT 46
US-08-332-766A-37/c
; Sequence 37, Application US/08332766A
; Patent No. 5843647
;  GENERAL INFORMATION:
;    APPLICANT:  JEFFREYS, Alec J.
;    APPLICANT:  ARMOUR, John
;    TITLE OF INVENTION:  SIMPLE TANDEM REPEATS
;    NUMBER OF SEQUENCES:  125
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D. C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/332,766A
;      FILING DATE:  01-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9326052.9
;      FILING DATE:  21-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BIRD, Donald J.
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  217211/M94/0434/GB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  124 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-332-766A-37

  Query Match             4.6%;  Score 46.4;  DB 2;  Length 124;
  Best Local Similarity   72.0%;  Pred. No. 0.00085;
  Matches   59;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy        233 GTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 292
              ||   || |||  || |||||| |||||||| ||| || ||||||||||| ||| || |
Db         82 GTGACGCCGCTNNCTTGCCTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 23

Qy        293 GCTTGCTTGCTCGCTGGCCTGC 314
              ||| ||| ||  ||  | | |
Db         22 TCTTTCTTTCTTTCTTTCTTTC 1


RESULT 47
US-09-128-155-16/c
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
;  APPLICANT: Pan, Yang
;  TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
;  TITLE OF INVENTION: AND USES THEREOF
;  FILE REFERENCE: 09404/052001
;  CURRENT APPLICATION NUMBER: US/09/128,155
;  CURRENT FILING DATE: 1998-08-03
;  EARLIER APPLICATION NUMBER: US 60/091,650
;  EARLIER FILING DATE: 1998-07-02
;  EARLIER APPLICATION NUMBER: US 60/054,646
;  EARLIER FILING DATE: 1997-08-04
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 16
;   LENGTH: 152331
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(152331)
;   OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

  Query Match             4.6%;  Score 46.2;  DB 3;  Length 152331;
  Best Local Similarity   59.6%;  Pred. No. 0.013;
  Matches   96;  Conservative    0;  Mismatches   63;  Indels    2;  Gaps    1;

Qy        243 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 302
              ||  |  | | ||| ||| ||  ||  | | ||| ||| ||  |  || |||||||| |
Db     147104 TCTTTTTCTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTTCCTGCTTGCTTGCTTTC 147045

Qy        303 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 362
              |  ||  ||| | | | | ||  ||  ||  ||| ||| |||  |      ||| ||| |
Db     147044 TTTCTTTCCTTCCTTCCTTCTTTCTCTCC--CTTTCTTTCTCTTTCTTTCTCTTTCTTTC 146987

Qy        363 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403
              |  || | | ||| ||| ||  ||  | | ||| ||| ||
Db     146986 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 146946


RESULT 48
US-09-128-155-17
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
;  APPLICANT: Pan, Yang
;  TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
;  TITLE OF INVENTION: AND USES THEREOF
;  FILE REFERENCE: 09404/052001
;  CURRENT APPLICATION NUMBER: US/09/128,155
;  CURRENT FILING DATE: 1998-08-03
;  EARLIER APPLICATION NUMBER: US 60/091,650
;  EARLIER FILING DATE: 1998-07-02
;  EARLIER APPLICATION NUMBER: US 60/054,646
;  EARLIER FILING DATE: 1997-08-04
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 3.0
```

```
; SEQ ID NO 17
;   LENGTH: 176373
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(176373)
;   OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

  Query Match             4.6%;  Score 46.2;  DB 3;  Length 176373;
  Best Local Similarity   59.6%;  Pred. No. 0.014;
  Matches   96;  Conservative    0;  Mismatches   63;  Indels    2;  Gaps    1;

Qy        243 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 302
              ||  |  | | ||| ||| ||  ||  | | ||| ||| ||  |  || |||||||||| |
Db     122289 TCTTTATCTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTTCCTGCTTGCTTGCTTTC 122348

Qy        303 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 362
              |  ||  ||| | | | | ||  ||  ||  ||| ||| |||  |       ||| ||| |
Db     122349 TTTCTTTCCTTCCTTCCTTCTTTCTCTCC--CTTTCTTTCTCTTTCTTTCTCTTTCTTTC 122406

Qy        363 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403
              |  ||  | | ||| ||| ||  ||  | | ||| ||| ||
Db     122407 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 122447


RESULT 49
US-09-007-005-32/c
; Sequence 32, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32
;   LENGTH: 248
;   TYPE: RNA
;   ORGANISM: Homo sapiens
US-09-007-005-32

  Query Match             4.5%;  Score 45.2;  DB 3;  Length 248;
  Best Local Similarity   19.2%;  Pred. No. 0.0023;
  Matches   38;  Conservative   86;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        219 CTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 278
              |:|:   : : :| |:    |: : :|:|  : :|:|:|:|: :|: :  |:|: : : :
Db        222 CYGYCYAYGYCYTYGYTYTYCYCYAYGYTYTYTYGYTYGYTYTYTYCYAYGYCYTYGYTY 163

Qy        279 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 338
              |: : : :   :|:|:|: :|: : :|:   : : : :|: : : : :|  : : : : :
Db        162 TYCYAYCYGYAYCYGYTYTYTYAYCYGYCYAYGYCYAYGYGYTYCYTYTYCYTYTYCYAY 103

Qy        339 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 398
               : : : :|  : :|:|: :|: :|:|:|  :|: :|: : :|: : :   : :|:|: :
Db        102 GYAYGYAYTYCYAYGYTYTYTYCYTYGYTYTYCYTYTYCYAYGYCYCYAYTYTYGYTYAY 43

Qy        399 TTGCTCGCTGGCCTGCTT 416
               : :|:|:|  : : :|:
Db         42 AYTYTYGYTYAYAYAYTY 25


RESULT 50
US-09-244-796-32/c
; Sequence 32, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350007
;  CURRENT APPLICATION NUMBER: US/09/244,796
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  EARLIER APPLICATION NUMBER: 09/007,005
;  EARLIER FILING DATE: 1998-01-14
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32
;   LENGTH: 248
;   TYPE: RNA
;   ORGANISM: Homo sapiens
US-09-244-796-32

  Query Match             4.5%;  Score 45.2;  DB 3;  Length 248;
  Best Local Similarity   19.2%;  Pred. No. 0.0023;
  Matches   38;  Conservative   86;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        219 CTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 278
              |:|:   : : :| |:    |: : :|:|  : :|:|:|:|: :|: :  |:|: : : :
Db        222 CYGYCYAYGYCYTYGYTYTYCYCYAYGYTYTYTYGYTYGYTYTYTYCYAYGYCYTYGYTY 163

Qy        279 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 338
              |: : : :   :|:|:|: :|: : :|:   : : : :|: : : : :|  : : : : :
Db        162 TYCYAYCYGYAYCYGYTYTYTYAYCYGYCYAYGYCYAYGYGYTYCYTYTYCYTYTYCYAY 103

Qy        339 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 398
               : : : :|  : :|:|: :|: :|:|:|  :|: :|: : :|: : :   : :|:|: :
Db        102 GYAYGYAYTYCYAYGYTYTYTYCYTYGYTYTYCYTYTYCYAYGYCYCYAYTYTYGYTYAY 43

Qy        399 TTGCTCGCTGGCCTGCTT 416
               : :|:|:|  : : :|:
Db         42 AYTYTYGYTYAYAYAYTY 25


RESULT 51
US-09-007-005-3/c
; Sequence 3, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 277
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
```

```
US-09-007-005-3

  Query Match             4.5%;  Score 45.2;  DB 3;  Length 277;
  Best Local Similarity   19.2%;  Pred. No. 0.0024;
  Matches   38;  Conservative   86;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        219 CTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 278
              |:|:   : : :| |:    |: : :|:|  : :|:|:|:|: :|: :  |:|: : : :
Db        222 CYGYCYAYGYCYTYGYTYTYCYCYAYGYTYTYTYGYTYGYTYTYTYCYAYGYCYTYGYTY 163

Qy        279 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 338
              |: : : :   :|:|:|: :|: : :|:   : : : :|: : : : :|  : : : : :
Db        162 TYCYAYCYGYAYCYGYTYTYTYAYCYGYCYAYGYCYAYGYGYTYCYTYTYCYTYTYCYAY 103

Qy        339 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 398
              : : : :|  : :|:|: :|: :|:|:|  :|: :|: : :|: : :   : :|:|: :
Db        102 GYAYGYAYTYCYAYGYTYTYTYCYTYGYTYTYCYTYTYCYAYGYCYCYAYTYTYGYTYAY 43

Qy        399 TTGCTCGCTGGCCTGCTT 416
              : :|:|:|  : : :|:
Db         42 AYTYTYGYTYAYAYAYTY 25


RESULT 52
US-09-244-796-3/c
; Sequence 3, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350007
;  CURRENT APPLICATION NUMBER: US/09/244,796
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  EARLIER APPLICATION NUMBER: 09/007,005
;  EARLIER FILING DATE: 1998-01-14
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 277
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
US-09-244-796-3

  Query Match             4.5%;  Score 45.2;  DB 3;  Length 277;
  Best Local Similarity   19.2%;  Pred. No. 0.0024;
  Matches   38;  Conservative   86;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        219 CTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 278
              |:|:   : : :| |:    |: : :|:|  : :|:|:|:|: :|: :  |:|: : : :
Db        222 CYGYCYAYGYCYTYGYTYTYCYCYAYGYTYTYTYGYTYGYTYTYTYCYAYGYCYTYGYTY 163

Qy        279 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 338
              |: : : :   :|:|:|: :|: : :|:   : : : :|: : : : :|  : : : : :
Db        162 TYCYAYCYGYAYCYGYTYTYTYAYCYGYCYAYGYCYAYGYGYTYCYTYTYCYTYTYCYAY 103

Qy        339 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 398
              : : : :|  : :|:|: :|: :|:|:|  :|: :|: : :|: : :   : :|:|: :
Db        102 GYAYGYAYTYCYAYGYTYTYTYCYTYGYTYTYCYTYTYCYAYGYCYCYAYTYTYGYTYAY 43

Qy        399 TTGCTCGCTGGCCTGCTT 416
              : :|:|:|  : : :|:
Db         42 AYTYTYGYTYAYAYAYTY 25


RESULT 53
US-07-753-520B-2/c
; Sequence 2, Application US/07753520B
; Patent No. 5352595
;  GENERAL INFORMATION:
;    APPLICANT:  Tapscott,J.;Weintraub,H.M.;Palmer,T.D.
;    TITLE OF INVENTION:  "MyoD REGULATORY REGION"
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;      STREET:  2800 Pacific First Center, 1420 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-5.25 inch, 1.2Mb storage
;      COMPUTER:  IBM PC/386 Compatible
;      OPERATING SYSTEM:  MS-DOS 4.01
;      SOFTWARE:  Word for Windows-t
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/753,520B
;      FILING DATE:  19910903
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  none
;      FILING DATE:  none
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sundsmo,John,S.
;      REGISTRATION NUMBER:  34,446
;      REFERENCE/DOCKET NUMBER:  FHCR-1-5789
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  1-206-682-8100; 1-206-224-0727 (direct)
;      TELEFAX:  1-206-224-0779
;      TELEX:  4938023
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1978 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;      DESCRIPTION:  KX fragment MyoD Enhancer; distal regulatory region myoD gene; F
US-07-753-520B-2

  Query Match             4.5%;  Score 45.2;  DB 1;  Length 1978;
  Best Local Similarity   54.1%;  Pred. No. 0.0049;
  Matches   92;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        259 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 318
              || ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| |
Db       1371 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTCCTTCC 1312

Qy        319 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 378
              || ||  ||  | | ||| ||  ||| ||  |   ||  ||  ||| ||      ||  |
Db       1311 TTCCTTCCTTCCTTCCTTCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCTTCCCTCTC 1252

Qy        379 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 428
              |  ||| ||  |   ||  ||  ||| ||   ||     | |||   |||
Db       1251 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTTATCTCAAACCCGAGCAAACA 1202


RESULT 54
US-07-753-520B-3/c
; Sequence 3, Application US/07753520B
; Patent No. 5352595
;  GENERAL INFORMATION:
;    APPLICANT:  Tapscott,J.;Weintraub,H.M.;Palmer,T.D.
;    TITLE OF INVENTION:  "MyoD REGULATORY REGION"
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;      STREET:  2800 Pacific First Center, 1420 Fifth Avenue
```

```
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98101-2347
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette-5.25 inch, 1.2Mb storage
;       COMPUTER:  IBM PC/386 Compatible
;       OPERATING SYSTEM:  MS-DOS 4.01
;       SOFTWARE:  Word for Windows-t
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/753,520B
;       FILING DATE:  19910903
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  none
;       FILING DATE:  none
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sundsmo,John,S.
;       REGISTRATION NUMBER:  34,446
;       REFERENCE/DOCKET NUMBER:  FHCR-1-5789
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  1-206-682-8100; 1-206-224-0727 (direct)
;       TELEFAX:  1-206-224-0779
;       TELEX:  4938023
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  9115 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Other;plasmid DNA
;       DESCRIPTION:  pLHDMDN-53: 5'LTR (position 1-1159); Y+ (position
;       DESCRIPTION:  1159-1640); HisD (position 1641-3007); Myo-D 531.4 ApaI fragment
;       DESCRIPTION:  (position 3008-5248); driving neo (position 5249-6117); with a
;       DESCRIPTION:  3'LTR (position 6118-6823) coupled to a pBR322 plasmid (position
;       DESCRIPTION:  6824-9115); Figures 7A-7D.
US-07-753-520B-3

  Query Match             4.5%;  Score 45.2;  DB 1;  Length 9115;
  Best Local Similarity   54.1%;  Pred. No. 0.0086;
  Matches   92;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        259 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 318
              || ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| |
Db       4378 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTCCTTCC 4319

Qy        319 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 378
              || ||  ||  | | ||| ||  ||| ||  |   ||  ||  ||| ||      ||  |
Db       4318 TTCCTTCCTTCCTTCCTTCCTCCCTCCCTCCCTCCCTCCCTCCCTCCTTCCCTCTC 4259

Qy        379 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 428
              |  ||| ||  |   ||  ||  ||| ||   ||     | |||   |||
Db       4258 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTTATCTCAAACCCGAGCAAACA 4209


RESULT 55
US-08-585-593A-42/c
; Sequence 42, Application US/08585593A
; Patent No. 6503706
;  GENERAL INFORMATION:
;    APPLICANT:  ABKEN, Hinrich J
;    APPLICANT:  ALBERT, Winfried
;    APPLICANT:  JUNGFER, Herbert
;    TITLE OF INVENTION:  METHOD OF IDENTIFYING HUMAN AND ANIMAL
;    TITLE OF INVENTION:  CELLS CAPABLE OF UNLIMITED PROLIFERATION OR TUMOR
;    TITLE OF INVENTION:  FORMATION
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nikaido, Marmelstein, Murray & Oram LLP
;      STREET:  655 Fifteenth Street N.W. Suite 330
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-5701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/585,593A
;      FILING DATE:  16-JAN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP94/02307
;      FILING DATE:  13-JUL-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P 43 23 727.4
;      FILING DATE:  15-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murray, Robert B.
;      REGISTRATION NUMBER:  22,980
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)638-5000
;      TELEFAX:  (202)638-4810
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  275 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-585-593A-42

  Query Match             4.5%;  Score 45;  DB 4;  Length 275;
  Best Local Similarity   56.4%;  Pred. No. 0.0027;
  Matches   84;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy        239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
              || ||   |   || |||  |  ||| ||  |   ||  ||  ||| ||  | | ||| |
Db        246 CTTCTTCTTCTTCTTCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTTTCTTTCTTTC 187

Qy        299 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 358
              || ||  ||  | | ||| ||  ||  ||  |   ||| ||| ||  ||  | | ||| |
Db        186 TTTCTTTCTTTCTTTCTTCCTGTCTGCCTACCTGCCTTCCTTCCTTCCTTCCTTCCTTCC 127

Qy        359 TTGCTCGCTGGCCTGCTTGCTTGCTCGCT 387
              || ||  ||  | | ||| ||| ||  ||
Db        126 TTCCTTCCTTCCTTCCTTCCTTCCTTTCT 98


RESULT 56
US-08-485-284A-1/c
; Sequence 1, Application US/08485284A
; Patent No. 5750372
;  GENERAL INFORMATION:
;    APPLICANT:  SAKAI, YASUYOSHI
;    APPLICANT:  TANI, YOSHIKI
;    APPLICANT:  SHIBANO, YUJI
;    APPLICANT:  KONDO, HIROTO
;    APPLICANT:  HATANAKA, HARUYO
;    TITLE OF INVENTION:  NOVEL VECTOR HAVING PROMOTER THAT IS
;    TITLE OF INVENTION:  INDUCIBLE BY METHANOL AND/OR GLYCEROL
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D. C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;  CURRENT APPLICATION DATA:
;    APPLICATION NUMBER:  US/08/485,284A
;    FILING DATE:  07-JUN-1995
;    CLASSIFICATION:  435
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER:  JP 43361/1992
;    FILING DATE:  28-FEB-1992
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER:  US 08/025,416
;    FILING DATE:  01-MAR-1993
;  ATTORNEY/AGENT INFORMATION:
;    NAME:  WHITE JR., PAUL E.
;    REGISTRATION NUMBER:  32,011
;    REFERENCE/DOCKET NUMBER:  217755/FPS38209US
;  TELECOMMUNICATION INFORMATION:
;    TELEPHONE:  (202) 861-3000
;    TELEFAX:  (202) 822-0944
;    TELEX:  6714627 CUSH
; INFORMATION FOR SEQ ID NO:  1:
;  SEQUENCE CHARACTERISTICS:
;    LENGTH:  1667 base pairs
;    TYPE:  nucleic acid
;    STRANDEDNESS:  double
;    TOPOLOGY:  unknown
;  MOLECULE TYPE:  genomic DNA
US-08-485-284A-1

  Query Match             4.4%;  Score 43.8;  DB 1;  Length 1667;
  Best Local Similarity   53.9%;  Pred. No. 0.011;
  Matches   90;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy        331 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 390
              |||| ||| ||| || |||||| ||| ||| || |||||| ||| || | |
Db        624 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGGCCCGTATT 565

Qy        391 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAG 450
               ||  ||  ||   |   | ||| |||       |  ||  | |   |  |  |||   |
Db        564 TTGGCTGGCTGGCTGACTGGCTGATTGAGAGATTGGGAGATGAGGCACACGGCTGCAGGG 505

Qy        451 GGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTG 497
                ||||  |  || ||   | ||| ||     || | ||| ||| ||
Db        504 TAGACGCAAGTCCAGCTGCACGGCGTGCATCCAGAGATGCGGGGGTG 458


RESULT 57
US-09-776-976-7/c
; Sequence 7, Application US/09776976
; Patent No. 6566332
; GENERAL INFORMATION:
;  APPLICANT: Fruebis, Joachim
;  APPLICANT:  Erickson, Mary Ruth
;  APPLICANT:  Yen, Frances
;  APPLICANT:  Bihain, Bernard
;  TITLE OF INVENTION: OBG3 Globular Head and Uses Thereof for Decreasing Body Mass
;  FILE REFERENCE: 76.US4.REG
;  CURRENT APPLICATION NUMBER: US/09/776,976
;  CURRENT FILING DATE:  2001-02-05
;  PRIOR APPLICATION NUMBER: US 09/758,055
;  PRIOR FILING DATE: 2001-01-10
;  PRIOR APPLICATION NUMBER: US 60/176,228
;  PRIOR FILING DATE: 2000-01-14
;  PRIOR APPLICATION NUMBER: US 60/198,087
;  PRIOR FILING DATE: 2000-04-13
;  PRIOR APPLICATION NUMBER: US 60/299,881
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: Patent.pm
; SEQ ID NO 7
;   LENGTH: 20966
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..4811
;   OTHER INFORMATION: 5' regulatory region
;   NAME/KEY: exon
;   LOCATION: 4812..4851
;   OTHER INFORMATION: exon 1
;   NAME/KEY: exon
;   LOCATION: 15144..15365
;   OTHER INFORMATION: exon 2
;   NAME/KEY: exon
;   LOCATION: 16277..20559
;   OTHER INFORMATION: exon 3
;   NAME/KEY: misc_feature
;   LOCATION: 20560..20966
;   OTHER INFORMATION: 3' regulatory region
;   NAME/KEY: allele
;   LOCATION: 3787
;   OTHER INFORMATION: 9-27-261 : polymorphic base G or C
;   NAME/KEY: allele
;   LOCATION: 11118
;   OTHER INFORMATION: 99-14387-129 : polymorphic base A or C
;   NAME/KEY: allele
;   LOCATION: 15120
;   OTHER INFORMATION: 9-12-48 : polymorphic base C or T
;   NAME/KEY: allele
;   LOCATION: 15196
;   OTHER INFORMATION: 9-12-124 : polymorphic base G or T
;   NAME/KEY: allele
;   LOCATION: 15427
;   OTHER INFORMATION: 9-12-355 : polymorphic base G or T
;   NAME/KEY: allele
;   LOCATION: 15500
;   OTHER INFORMATION: 9-12-428 : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 15863
;   OTHER INFORMATION: 99-14405-105 : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 17170
;   OTHER INFORMATION: 9-16-189 : polymorphic base deletion of A
;   NAME/KEY: primer_bind
;   LOCATION: 3528..3545
;   OTHER INFORMATION: 9-27.pu
;   NAME/KEY: primer_bind
;   LOCATION: 3928..3946
;   OTHER INFORMATION: 9-27.rp complement
;   NAME/KEY: primer_bind
;   LOCATION: 10990..11008
;   OTHER INFORMATION: 99-14387.pu
;   NAME/KEY: primer_bind
;   LOCATION: 11423..11442
;   OTHER INFORMATION: 99-14387.rp complement
;   NAME/KEY: primer_bind
;   LOCATION: 15073..15092
;   OTHER INFORMATION: 9-12.pu
;   NAME/KEY: primer_bind
;   LOCATION: 15503..15520
;   OTHER INFORMATION: 9-12.rp complement
;   NAME/KEY: primer_bind
;   LOCATION: 15759..15776
;   OTHER INFORMATION: 99-14405.pu
;   NAME/KEY: primer_bind
;   LOCATION: 16191..16211
;   OTHER INFORMATION: 99-14405.rp complement
;   NAME/KEY: primer_bind
;   LOCATION: 16982..17001
;   OTHER INFORMATION: 9-16.pu
;   NAME/KEY: primer_bind
;   LOCATION: 17384..17402
;   OTHER INFORMATION: 9-16.rp complement
;   NAME/KEY: misc_binding
;   LOCATION: 3775..3799
;   OTHER INFORMATION: 9-27-261.probe
;   NAME/KEY: misc_binding
;   LOCATION: 11106..11130
```

```
;   OTHER INFORMATION: 99-14387-129.probe
;   NAME/KEY: misc_binding
;   LOCATION: 15108..15132
;   OTHER INFORMATION: 9-12-48.probe
;   NAME/KEY: misc_binding
;   LOCATION: 15184..15208
;   OTHER INFORMATION: 9-12-124.probe
;   NAME/KEY: misc_binding
;   LOCATION: 15415..15439
;   OTHER INFORMATION: 9-12-355.probe
;   NAME/KEY: misc_binding
;   LOCATION: 15488..15512
;   OTHER INFORMATION: 9-12-428.probe
;   NAME/KEY: misc_binding
;   LOCATION: 15851..15875
;   OTHER INFORMATION: 99-14405-105.probe
;   NAME/KEY: misc_binding
;   LOCATION: 17158..17182
;   OTHER INFORMATION: 9-16-189.probe
;   NAME/KEY: primer_bind
;   LOCATION: 3768..3786
;   OTHER INFORMATION: 9-27-261.mis
;   NAME/KEY: primer_bind
;   LOCATION: 3788..3806
;   OTHER INFORMATION: 9-27-261.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 11099..11117
;   OTHER INFORMATION: 99-14387-129.mis
;   NAME/KEY: primer_bind
;   LOCATION: 11119..11137
;   OTHER INFORMATION: 99-14387-129.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 15101..15119
;   OTHER INFORMATION: 9-12-48.mis
;   NAME/KEY: primer_bind
;   LOCATION: 15121..15139
;   OTHER INFORMATION: 9-12-48.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 15177..15195
;   OTHER INFORMATION: 9-12-124.mis
;   NAME/KEY: primer_bind
;   LOCATION: 15197..15215
;   OTHER INFORMATION: 9-12-124.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 15408..15426
;   OTHER INFORMATION: 9-12-355.mis
;   NAME/KEY: primer_bind
;   LOCATION: 15428..15446
;   OTHER INFORMATION: 9-12-355.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 15481..15499
;   OTHER INFORMATION: 9-12-428.mis
;   NAME/KEY: primer_bind
;   LOCATION: 15501..15519
;   OTHER INFORMATION: 9-12-428.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 15844..15862
;   OTHER INFORMATION: 99-14405-105.mis
;   NAME/KEY: primer_bind
;   LOCATION: 15864..15882
;   OTHER INFORMATION: 99-14405-105.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 17151..17169
;   OTHER INFORMATION: 9-16-189.mis
;   NAME/KEY: primer_bind
;   LOCATION: 17171..17189
;   OTHER INFORMATION: 9-16-189.mis complement
US-09-776-976-7

  Query Match             4.4%;  Score 43.8;  DB 4;  Length 20966;
  Best Local Similarity   55.6%;  Pred. No. 0.028;
  Matches   84;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        250 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 309
              | | ||| ||| ||  ||  ||  | |  ||·||| ||  ||    | ||| ||| ||  |
Db      10425 CTTCCTTCCTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCTTTC 10366

Qy        310 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 369
              | | ||| ||| ||  ||    | ||| ||| ||  ||    | ||| ||| ||  ||
Db      10365 CTTCCTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCCTTC 10306

Qy        370 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
              | | |||  || ||| ||  |    ||  ||
Db      10305 CTTCCTTCTTTTCTCTCTCTCTCTTTTTTTT 10275


RESULT 58
US-09-909-547-7/c
; Sequence 7, Application US/09909547
; Patent No. 6579852
; GENERAL INFORMATION:
;  APPLICANT: Fruebis, Joachim
;  APPLICANT:  Erickson, Mary Ruth
;  APPLICANT:  Yen, Frances
;  APPLICANT:  Bihain, Bernard
;  TITLE OF INVENTION: OBG3 Globular Head and Uses Thereof for Decreasing Body Mass
;  FILE REFERENCE: 76.US6.CIP
;  CURRENT APPLICATION NUMBER: US/09/909,547
;  CURRENT FILING DATE:  2001-07-19
;  PRIOR APPLICATION NUMBER: US 09/776,976
;  PRIOR FILING DATE: 2001-02-05
;  PRIOR APPLICATION NUMBER: US 09/758,055
;  PRIOR FILING DATE: 2001-01-10
;  PRIOR APPLICATION NUMBER: US 60/299,881
;  PRIOR FILING DATE: 2000-09-01
;  PRIOR APPLICATION NUMBER: US 60/198,087
;  PRIOR FILING DATE: 2000-04-13
;  PRIOR APPLICATION NUMBER: US 60/176,228
;  PRIOR FILING DATE: 2000-01-14
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: Patent.pm
; SEQ ID NO 7
;   LENGTH: 20966
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..4811
;   OTHER INFORMATION: 5' regulatory region
;   NAME/KEY: exon
;   LOCATION: 4812..4851
;   OTHER INFORMATION: exon 1
;   NAME/KEY: exon
;   LOCATION: 15144..15365
;   OTHER INFORMATION: exon 2
;   NAME/KEY: exon
;   LOCATION: 16277..20559
;   OTHER INFORMATION: exon 3
;   NAME/KEY: misc_feature
;   LOCATION: 20560..20966
;   OTHER INFORMATION: 3' regulatory region
;   NAME/KEY: allele
;   LOCATION: 3787
;   OTHER INFORMATION: 9-27-261 : polymorphic base G or C
;   NAME/KEY: allele
;   LOCATION: 11118
;   OTHER INFORMATION: 99-14387-129 : polymorphic base A or C
;   NAME/KEY: allele
;   LOCATION: 15120
;   OTHER INFORMATION: 9-12-48 : polymorphic base C or T
;   NAME/KEY: allele
;   LOCATION: 15196
;   OTHER INFORMATION: 9-12-124 : polymorphic base G or T
;   NAME/KEY: allele
;   LOCATION: 15427
;   OTHER INFORMATION: 9-12-355 : polymorphic base G or T
```

```
;   NAME/KEY: allele
;   LOCATION: 15500
;   OTHER INFORMATION: 9-12-428 : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 15863
;   OTHER INFORMATION: 99-14405-105 : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 17170
;   OTHER INFORMATION: 9-16-189 : polymorphic base deletion of A
;   NAME/KEY: primer_bind
;   LOCATION: 3528..3545
;   OTHER INFORMATION: 9-27.pu
;   NAME/KEY: primer_bind
;   LOCATION: 3928..3946
;   OTHER INFORMATION: 9-27.rp complement
;   NAME/KEY: primer_bind
;   LOCATION: 10990..11008
;   OTHER INFORMATION: 99-14387.pu
;   NAME/KEY: primer_bind
;   LOCATION: 11423..11442
;   OTHER INFORMATION: 99-14387.rp complement
;   NAME/KEY: primer_bind
;   LOCATION: 15073..15092
;   OTHER INFORMATION: 9-12.pu
;   NAME/KEY: primer_bind
;   LOCATION: 15503..15520
;   OTHER INFORMATION: 9-12.rp complement
;   NAME/KEY: primer_bind
;   LOCATION: 15759..15776
;   OTHER INFORMATION: 99-14405.pu
;   NAME/KEY: primer_bind
;   LOCATION: 16191..16211
;   OTHER INFORMATION: 99-14405.rp complement
;   NAME/KEY: primer_bind
;   LOCATION: 16982..17001
;   OTHER INFORMATION: 9-16.pu
;   NAME/KEY: primer_bind
;   LOCATION: 17384..17402
;   OTHER INFORMATION: 9-16.rp complement
;   NAME/KEY: misc_binding
;   LOCATION: 3775..3799
;   OTHER INFORMATION: 9-27-261.probe
;   NAME/KEY: misc_binding
;   LOCATION: 11106..11130
;   OTHER INFORMATION: 99-14387-129.probe
;   NAME/KEY: misc_binding
;   LOCATION: 15108..15132
;   OTHER INFORMATION: 9-12-48.probe
;   NAME/KEY: misc_binding
;   LOCATION: 15184..15208
;   OTHER INFORMATION: 9-12-124.probe
;   NAME/KEY: misc_binding
;   LOCATION: 15415..15439
;   OTHER INFORMATION: 9-12-355.probe
;   NAME/KEY: misc_binding
;   LOCATION: 15488..15512
;   OTHER INFORMATION: 9-12-428.probe
;   NAME/KEY: misc_binding
;   LOCATION: 15851..15875
;   OTHER INFORMATION: 99-14405-105.probe
;   NAME/KEY: misc_binding
;   LOCATION: 17158..17182
;   OTHER INFORMATION: 9-16-189.probe
;   NAME/KEY: primer_bind
;   LOCATION: 3768..3786
;   OTHER INFORMATION: 9-27-261.mis
;   NAME/KEY: primer_bind
;   LOCATION: 3788..3806
;   OTHER INFORMATION: 9-27-261.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 11099..11117
;   OTHER INFORMATION: 99-14387-129.mis
;   NAME/KEY: primer_bind
;   LOCATION: 11119..11137
;   OTHER INFORMATION: 99-14387-129.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 15101..15119
;   OTHER INFORMATION: 9-12-48.mis
;   NAME/KEY: primer_bind
;   LOCATION: 15121..15139
;   OTHER INFORMATION: 9-12-48.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 15177..15195
;   OTHER INFORMATION: 9-12-124.mis
;   NAME/KEY: primer_bind
;   LOCATION: 15197..15215
;   OTHER INFORMATION: 9-12-124.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 15408..15426
;   OTHER INFORMATION: 9-12-355.mis
;   NAME/KEY: primer_bind
;   LOCATION: 15428..15446
;   OTHER INFORMATION: 9-12-355.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 15481..15499
;   OTHER INFORMATION: 9-12-428.mis
;   NAME/KEY: primer_bind
;   LOCATION: 15501..15519
;   OTHER INFORMATION: 9-12-428.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 15844..15862
;   OTHER INFORMATION: 99-14405-105.mis
;   NAME/KEY: primer_bind
;   LOCATION: 15864..15882
;   OTHER INFORMATION: 99-14405-105.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 17151..17169
;   OTHER INFORMATION: 9-16-189.mis
;   NAME/KEY: primer_bind
;   LOCATION: 17171..17189
;   OTHER INFORMATION: 9-16-189.mis complement
US-09-909-547-7

  Query Match             4.4%;  Score 43.8;  DB 4;  Length 20966;
  Best Local Similarity   55.6%;  Pred. No. 0.028;
  Matches   84;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        250 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 309
              | | ||| ||| ||  ||  | |  || ||| ||  ||    | ||| ||| ||   |
Db      10425 CTTCCTTCCTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCTTTC 10366

Qy        310 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 369
              | | ||| ||| ||  ||    | ||| ||| ||  ||    | ||| ||| ||  ||
Db      10365 CTTCCTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCCTTC 10306

Qy        370 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
              | | |||  || ||| ||  |    ||  ||
Db      10305 CTTCCTTCTTTTCTCTCTCTCTCTTTTTTTT 10275


RESULT 59
US-09-569-852B-1/c
; Sequence 1, Application US/09569852B
; Patent No. 6582909
; GENERAL INFORMATION:
;  APPLICANT: Bougueleret, Lydie
;  APPLICANT:  Bihain, Bernard
;  APPLICANT:  Denison, Blake
;  APPLICANT:  Yen-Potin, Frances
;  TITLE OF INVENTION: APM1 Biallelic Markers and Uses Thereof
;  FILE REFERENCE: GEN-T113XC2
;  CURRENT APPLICATION NUMBER: US/09/569,852B
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: PCT/IB99/01858
;  PRIOR FILING DATE: 1999-11-04
;  PRIOR APPLICATION NUMBER: US 09/434,848
```

; PRIOR FILING DATE: 1999-11-04
; PRIOR APPLICATION NUMBER: US 60/119,593
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/107,113
; PRIOR FILING DATE: 1998-11-04
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 20966
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(4811)
; OTHER INFORMATION: 5' regulatory region
; NAME/KEY: primer_bind
; LOCATION: (14683)..(14701)
; OTHER INFORMATION: 17-34-860.mis
; NAME/KEY: primer_bind
; LOCATION: (14703)..(14721)
; OTHER INFORMATION: 17-34-860.mis complement
; NAME/KEY: primer_bind
; LOCATION: (14738)..(14756)
; OTHER INFORMATION: 17-34-915.mis
; NAME/KEY: primer_bind
; LOCATION: (14758)..(14776)
; OTHER INFORMATION: 17-34-915.mis complement
; NAME/KEY: primer_bind
; LOCATION: (14796)..(14814)
; OTHER INFORMATION: 17-35-71.mis
; NAME/KEY: primer_bind
; LOCATION: (14816)..(14834)
; OTHER INFORMATION: 17-35-71.mis complement
; NAME/KEY: primer_bind
; LOCATION: (15031)..(15049)
; OTHER INFORMATION: 17-35-306.mis
; NAME/KEY: primer_bind
; LOCATION: (15051)..(15069)
; OTHER INFORMATION: 17-35-306.mis complement
; NAME/KEY: primer_bind
; LOCATION: (15101)..(15119)
; OTHER INFORMATION: 9-12-48.mis
; NAME/KEY: primer_bind
; LOCATION: (15121)..(15139)
; OTHER INFORMATION: 9-12-48.mis complement
; NAME/KEY: primer_bind
; LOCATION: (15177)..(15195)
; OTHER INFORMATION: 9-12-124.mis
; NAME/KEY: primer_bind
; LOCATION: (15197)..(15215)
; OTHER INFORMATION: 9-12-124.mis complement
; NAME/KEY: primer_bind
; LOCATION: (15408)..(15426)
; OTHER INFORMATION: 9-12-355.mis
; NAME/KEY: primer_bind
; LOCATION: (15428)..(15446)
; OTHER INFORMATION: 9-12-355.mis complement
; NAME/KEY: primer_bind
; LOCATION: (15481)..(15499)
; OTHER INFORMATION: 9-12-428.mis
; NAME/KEY: primer_bind
; LOCATION: (15501)..(15519)
; OTHER INFORMATION: 9-12-428.mis complement
; NAME/KEY: primer_bind
; LOCATION: (15661)..(15679)
; OTHER INFORMATION: 17-36-47.mis
; NAME/KEY: primer_bind
; LOCATION: (15681)..(15699)
; OTHER INFORMATION: 17-36-47.mis complement
; NAME/KEY: primer_bind
; LOCATION: (15771)..(15789)
; OTHER INFORMATION: 17-36-120.mis
; NAME/KEY: primer_bind
; LOCATION: (15791)..(15809)
; OTHER INFORMATION: 17-36-120.mis complement
; NAME/KEY: primer_bind
; LOCATION: (15844)..(15862)
; OTHER INFORMATION: 99-14405-105.mis
; NAME/KEY: primer_bind
; LOCATION: (15864)..(15882)
; OTHER INFORMATION: 99-14405-105.mis complement
; NAME/KEY: primer_bind
; LOCATION: (17151)..(17169)
; OTHER INFORMATION: 9-16-189.mis
; NAME/KEY: primer_bind
; LOCATION: (17171)..(17189)
; OTHER INFORMATION: 9-16-189.mis complement
; NAME/KEY: primer_bind
; LOCATION: (17810)..(17828)
; OTHER INFORMATION: 17-37-629.mis
; NAME/KEY: primer_bind
; LOCATION: (17830)..(17848)
; OTHER INFORMATION: 17-37-629.mis complement
; NAME/KEY: primer_bind
; LOCATION: (17992)..(18010)
; OTHER INFORMATION: 17-37-811.mis
; NAME/KEY: primer_bind
; LOCATION: (18012)..(18030)
; OTHER INFORMATION: 17-37-811.mis complement
; NAME/KEY: primer_bind
; LOCATION: (18470)..(18488)
; OTHER INFORMATION: 17-38-349.mis
; NAME/KEY: primer_bind
; LOCATION: (18490)..(18508)
; OTHER INFORMATION: 17-38-349.mis complement
; NAME/KEY: primer_bind
; LOCATION: (926)..(944)
; OTHER INFORMATION: 17-30-216.mis
; NAME/KEY: primer_bind
; LOCATION: (946)..(964)
; OTHER INFORMATION: 17-30-216.mis complement
; NAME/KEY: primer_bind
; LOCATION: (3719)..(3737)
; OTHER INFORMATION: 9-27-211.mis
; NAME/KEY: primer_bind
; LOCATION: (3739)..(3757)
; OTHER INFORMATION: 9-27-211.mis complement
; NAME/KEY: primer_bind
; LOCATION: (3754)..(3772)
; OTHER INFORMATION: 9-27-246.mis
; NAME/KEY: primer_bind
; LOCATION: (3774)..(3792)
; OTHER INFORMATION: 9-27-246.mis complement
; NAME/KEY: primer_bind
; LOCATION: (3768)..(3786)
; OTHER INFORMATION: 9-27-261.mis
; NAME/KEY: primer_bind
; LOCATION: (3788)..(3806)
; OTHER INFORMATION: 9-27-261.mis complement
; NAME/KEY: primer_bind
; LOCATION: (5076)..(5094)
; OTHER INFORMATION: 17-31-298.mis
; NAME/KEY: primer_bind
; LOCATION: (5096)..(5114)
; OTHER INFORMATION: 17-31-298.mis complement
; NAME/KEY: primer_bind
; LOCATION: (5191)..(5209)
; OTHER INFORMATION: 17-31-413.mis
; NAME/KEY: primer_bind
; LOCATION: (5211)..(5229)
; OTHER INFORMATION: 17-31-413.mis complement
; NAME/KEY: primer_bind
; LOCATION: (5364)..(5385)
; OTHER INFORMATION: 17-31.rp complement
; NAME/KEY: primer_bind
; LOCATION: (10618)..(10636)

```
;  'OTHER INFORMATION: 17-32-24.mis
;   NAME/KEY: primer_bind
;   LOCATION: (10638)..(10656)
;   OTHER INFORMATION: 17-32-24.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (11020)..(11038)
;   OTHER INFORMATION: 99-14387-50.mis
;   NAME/KEY: primer_bind
;   LOCATION: (11040)..(11058)
;   OTHER INFORMATION: 99-14387-50.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (11099)..(11117)
;   OTHER INFORMATION: 99-14387-129.mis
;   NAME/KEY: primer_bind
;   LOCATION: (11119)..(11137)
;   OTHER INFORMATION: 99-14387-129.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (11169)..(11187)
;   OTHER INFORMATION: 99-14387-199.mis
;   NAME/KEY: primer_bind
;   LOCATION: (11189)..(11207)
;   OTHER INFORMATION: 99-14387-199.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (13954)..(13972)
;   OTHER INFORMATION: 17-33-TGAGACT.mis
;   NAME/KEY: primer_bind
;   LOCATION: (13974)..(13992)
;   OTHER INFORMATION: 17-33-TGAGACT.mis complement
;   NAME/KEY: exon
;   LOCATION: (4812)..(4851)
;   OTHER INFORMATION:
;   NAME/KEY: exon
;   LOCATION: (15144)..(15365)
;   OTHER INFORMATION:
;   NAME/KEY: exon
;   LOCATION: (16277)..(20559)
;   OTHER INFORMATION:
;   NAME/KEY: misc_feature
;   LOCATION: (20560)..(20966)
;   OTHER INFORMATION: 3' regulatory region

  Query Match             4.4%;  Score 43.8;  DB 4;  Length 20966;
  Best Local Similarity   55.6%;  Pred. No. 0.028;
  Matches   84;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        250 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 309
              | | ||| ||| ||  ||  | |  || ||| ||  ||    | ||| ||| ||   |
Db      10425 CTTCCTTCCTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCTTTC 10366

Qy        310 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 369
              | | ||| ||| ||  ||    | ||| ||| ||  ||    | ||| ||| ||  ||
Db      10365 CTTCCTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCCTTC 10306

Qy        370 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
              | | |||  || ||| ||  |    ||  ||
Db      10305 CTTCCTTCTTTTCTCTCTCTCTCTTTTTTTT 10275


RESULT 60
US-08-845-258-7
; Sequence 7, Application US/08845258
; Patent No. 6183976
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Lodes, Michael J.
;    APPLICANT:  Houghton, Raymond
;    APPLICANT:  Sleath, Paul R.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE DIAGNOSIS
;    TITLE OF INVENTION:  AND TREATMENT OF B. MICROTI INFECTION
;    NUMBER OF SEQUENCES:  53
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED AND BERRY
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/845,258
;      FILING DATE:  24-APR-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.426C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206)682-6031
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4223 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-845-258-7

  Query Match             4.3%;  Score 42.8;  DB 3;  Length 4223;
  Best Local Similarity   54.4%;  Pred. No. 0.029;
  Matches   86;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        268 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 327
              ||| ||| ||| |||  ||  || ||| ||| |||  |   || ||  ||| |||  |
Db       3568 GGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGTGTGCGTGTGTGCGTGCGTGTG 3627

Qy        328 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 387
              || ||  ||| |||  ||  |  ||| ||| |||  |   || ||  ||  ||   |
Db       3628 TGCGTGTGTGCGTGCGTGCGTGTGTGCGTGCGTGCGTGTGTGCGTGTGTGTGTGTGTG 3687

Qy        388 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 425
              |  ||  ||  ||   |   |  ||  |   || |||
Db       3688 TGTGTGTGTGTGTGTGTGTGTGTGTGTTGGGATGGG 3725


RESULT 61
US-08-845-258-45/c
; Sequence 45, Application US/08845258
; Patent No. 6183976
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Lodes, Michael J.
;    APPLICANT:  Houghton, Raymond
;    APPLICANT:  Sleath, Paul R.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE DIAGNOSIS
;    TITLE OF INVENTION:  AND TREATMENT OF B. MICROTI INFECTION
;    NUMBER OF SEQUENCES:  53
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED AND BERRY
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/845,258
;      FILING DATE:  24-APR-1997
;      CLASSIFICATION:  435
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.426C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206)682-6031
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4223 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-845-258-45

  Query Match             4.3%;  Score 42.8;  DB 3;  Length 4223;
  Best Local Similarity   54.4%;  Pred. No. 0.029;
  Matches   86;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        268 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 327
              ||| ||| ||| |||  ||  || ||| ||| |||  |   || ||  ||| |||  |
Db        656 GGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGTGTGCGTGTGTGCGTGCGTGTG 597

Qy        328 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 387
               || ||  ||| |||  ||  |  ||| ||| |||  |   || ||  ||  ||   |
Db        596 TGCGTGTGTGCGTGCGTGCGTGTGTGCGTGCGTGCGTGTGTGCGTGTGTGTGTGTGTG 537

Qy        388 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 425
              |  ||  ||  ||   |   |  ||  |   || |||
Db        536 TGTGTGTGTGTGTGTGTGTGTGTGTGTTGGGATGGG 499


RESULT 62
US-08-990-571-7
; Sequence 7, Application US/08990571
; Patent No. 6214971
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G. et al.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B. M
;    NUMBER OF SEQUENCES:  79
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED AND BERRY
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/990,571
;      FILING DATE:  11-DEC-1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.426C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206)682-6031
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4223 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-990-571-7

  Query Match             4.3%;  Score 42.8;  DB 3;  Length 4223;
  Best Local Similarity   54.4%;  Pred. No. 0.029;
  Matches   86;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        268 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 327
              ||| ||| ||| |||  ||  || ||| ||| |||  |   || ||  ||| |||  |
Db       3568 GGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGTGTGCGTGTGTGCGTGCGTGTG 3627

Qy        328 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 387
               || ||  ||| |||  ||  |  ||| ||| |||  |   || ||  ||  ||   |
Db       3628 TGCGTGTGTGCGTGCGTGCGTGTGTGCGTGCGTGCGTGTGTGCGTGTGTGTGTGTGTG 3687

Qy        388 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 425
              |  ||  ||  ||   |   |  ||  |   || |||
Db       3688 TGTGTGTGTGTGTGTGTGTGTGTGTGTTGGGATGGG 3725


RESULT 63
US-08-990-571-45/c
; Sequence 45, Application US/08990571
; Patent No. 6214971
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G. et al.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B.
;    NUMBER OF SEQUENCES:  79
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED AND BERRY
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/990,571
;      FILING DATE:  11-DEC-1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.426C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206)682-6031
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4223 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-990-571-45

  Query Match             4.3%;  Score 42.8;  DB 3;  Length 4223;
  Best Local Similarity   54.4%;  Pred. No. 0.029;
  Matches   86;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        268 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 327
              ||| ||| ||| |||  ||  || ||| ||| |||  |   || ||  ||| |||  |
Db        656 GGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGTGTGCGTGTGTGCGTGCGTGTG 597

Qy        328 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 387
               || ||  ||| |||  ||  |  ||| ||| |||  |   || ||  ||  ||   |
Db        596 TGCGTGTGTGCGTGCGTGCGTGTGTGCGTGCGTGCGTGTGTGCGTGTGTGTGTGTGTG 537

Qy        388 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 425
              |  ||  ||  ||   |   |  ||  |   || |||
Db        536 TGTGTGTGTGTGTGTGTGTGTGTGTGTTGGGATGGG 499


RESULT 64
US-08-723-142A-7
```

```
; Sequence 7, Application US/08723142A
; Patent No. 6306396
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Lodes, Michael J.
;    APPLICANT:  Houghton, Raymond
;    APPLICANT:  Sleath, Paul R.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE DIAGNOSIS
;    TITLE OF INVENTION:  AND TREATMENT OF B. MICROTI INFECTION
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED AND BERRY
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/723,142A
;      FILING DATE:  01-OCT-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.426
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206)682-6031
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4223 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-723-142A-7

  Query Match             4.3%;  Score 42.8;  DB 4;  Length 4223;
  Best Local Similarity   54.4%;  Pred. No. 0.029;
  Matches   86;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        268 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 327
              ||| ||| ||| ||| ||  || ||| ||| ||| |   || ||  ||| ||| |
Db       3568 GGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGTGTGCGTGTGTGCGTGCGTGTG 3627

Qy        328 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 387
              || ||  ||| ||| ||  |  ||| ||| ||| |   || ||  ||  ||   |
Db       3628 TGCGTGTGTGCGTGCGTGCGTGTGTGCGTGCGTGCGTGTGTGCGTGTGTGTGTGTGTG 3687

Qy        388 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 425
              |  ||  ||  ||   |   |  ||  |   || |||
Db       3688 TGTGTGTGTGTGTGTGTGTGTGTGTGTTGGGATGGG 3725


RESULT 65
US-08-723-142A-45/c
; Sequence 45, Application US/08723142A
; Patent No. 6306396
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Lodes, Michael J.
;    APPLICANT:  Houghton, Raymond
;    APPLICANT:  Sleath, Paul R.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE DIAGNOSIS
;    TITLE OF INVENTION:  AND TREATMENT OF B. MICROTI INFECTION
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED AND BERRY
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/723,142A
;      FILING DATE:  01-OCT-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.426
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206)682-6031
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4223 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-723-142A-45

  Query Match             4.3%;  Score 42.8;  DB 4;  Length 4223;
  Best Local Similarity   54.4%;  Pred. No. 0.029;
  Matches   86;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        268 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 327
              ||| ||| ||| ||| ||  || ||| ||| ||| |   || ||  ||| ||| |
Db        656 GGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGTGTGCGTGTGTGCGTGCGTGTG 597

Qy        328 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 387
              || ||  ||| ||| ||  |  ||| ||| ||| |   || ||  ||  ||   |
Db        596 TGCGTGTGTGCGTGCGTGCGTGTGTGCGTGCGTGCGTGTGTGCGTGTGTGTGTGTGTG 537

Qy        388 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 425
              |  ||  ||  ||   |   |  ||  |   || |||
Db        536 TGTGTGTGTGTGTGTGTGTGTGTGTGTTGGGATGGG 499


RESULT 66
US-09-528-784A-7
; Sequence 7, Application US/09528784A
; Patent No. 6451315
; GENERAL INFORMATION:
;  APPLICANT: Reed, Steven G.
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Houghton, Raymond L.
;  APPLICANT:  Sleath, Paul R.
;  APPLICANT:  McNeill, Patricia D.
;  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
;  TITLE OF INVENTION:  AND TREATMENT OF B. MICROTI INFECTION
;  FILE REFERENCE: 210121.426C4
;  CURRENT APPLICATION NUMBER: US/09/528,784A
;  CURRENT FILING DATE:  2000-03-17
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 4223
;   TYPE: DNA
;   ORGANISM: Babesia microti
US-09-528-784A-7

  Query Match             4.3%;  Score 42.8;  DB 4;  Length 4223;
  Best Local Similarity   54.4%;  Pred. No. 0.029;
  Matches   86;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        268 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 327
```

```
              ||| ||| ||| |||  ||  || ||| ||| |||  |   || ||  ||| |||  |
Db       3568 GGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGTGTGCGTGTGTGCGTGCGTGTG 3627

Qy        328 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 387
              || ||  ||| |||  ||  |  ||| ||| |||  |   || ||  ||  ||   |
Db       3628 TGCGTGTGTGCGTGCGTGCGTGTGTGCGTGCGTGCGTGTGTGCGTGTGTGTGTGTGTG 3687

Qy        388 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 425
              |  ||  ||  ||   |   |  ||  |   || |||
Db       3688 TGTGTGTGTGTGTGTGTGTGTGTGTGTTGGGATGGG 3725


RESULT 67
US-09-528-784A-45/c
; Sequence 45, Application US/09528784A
; Patent No. 6451315
; GENERAL INFORMATION:
;  APPLICANT: Reed, Steven G.
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Houghton, Raymond L.
;  APPLICANT:  Sleath, Paul R.
;  APPLICANT:  McNeill, Patricia D.
;  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
;  TITLE OF INVENTION:  AND TREATMENT OF B. MICROTI INFECTION
;  FILE REFERENCE: 210121.426C4
;  CURRENT APPLICATION NUMBER: US/09/528,784A
;  CURRENT FILING DATE:  2000-03-17
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 45
;   LENGTH: 4223
;   TYPE: DNA
;   ORGANISM: Babesia microti
US-09-528-784A-45

  Query Match             4.3%;  Score 42.8;  DB 4;  Length 4223;
  Best Local Similarity   54.4%;  Pred. No. 0.029;
  Matches   86;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        268 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 327
              ||| ||| ||| |||  ||  || ||| ||| |||  |   || ||  ||| |||  |
Db        656 GGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGTGTGCGTGTGTGCGTGCGTGTG 597

Qy        328 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 387
              || ||  ||| |||  ||  |  ||| ||| |||  |   || ||  ||  ||   |
Db        596 TGCGTGTGTGCGTGCGTGCGTGTGTGCGTGCGTGCGTGTGTGCGTGTGTGTGTGTGTG 537

Qy        388 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 425
              |  ||  ||  ||   |   |  ||  |   || |||
Db        536 TGTGTGTGTGTGTGTGTGTGTGTGTGTTGGGATGGG 499


RESULT 68
US-09-569-098A-7
; Sequence 7, Application US/09569098A
; Patent No. 6569433
; GENERAL INFORMATION:
;  APPLICANT: Reed, Steven G.
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Houghton, Raymond L.
;  APPLICANT:  Sleath, Paul R.
;  APPLICANT:  McNeill, Patricia D.
;  APPLICANT:  Homer, Mary
;  APPLICANT:  Secrist, Heather
;  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
;  TITLE OF INVENTION:  AND TREATMENT OF B. MICROTI INFECTION
;  FILE REFERENCE: 210121.426C5
;  CURRENT APPLICATION NUMBER: US/09/569,098A
;  CURRENT FILING DATE:  2000-05-10
;  NUMBER OF SEQ ID NOS: 110
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 4223
;   TYPE: DNA
;   ORGANISM: Babesia microti
US-09-569-098A-7

  Query Match             4.3%;  Score 42.8;  DB 4;  Length 4223;
  Best Local Similarity   54.4%;  Pred. No. 0.029;
  Matches   86;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        268 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 327
              ||| ||| ||| |||  ||  || ||| ||| |||  |   || ||  ||| |||  |
Db       3568 GGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGTGTGCGTGTGTGCGTGCGTGTG 3627

Qy        328 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 387
              || ||  ||| |||  ||  |  ||| ||| |||  |   || ||  ||  ||   |
Db       3628 TGCGTGTGTGCGTGCGTGCGTGTGTGCGTGCGTGCGTGTGTGCGTGTGTGTGTGTGTG 3687

Qy        388 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 425
              |  ||  ||  ||   |   |  ||  |   || |||
Db       3688 TGTGTGTGTGTGTGTGTGTGTGTGTGTTGGGATGGG 3725


RESULT 69
US-09-569-098A-45/c
; Sequence 45, Application US/09569098A
; Patent No. 6569433
; GENERAL INFORMATION:
;  APPLICANT: Reed, Steven G.
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Houghton, Raymond L.
;  APPLICANT:  Sleath, Paul R.
;  APPLICANT:  McNeill, Patricia D.
;  APPLICANT:  Homer, Mary
;  APPLICANT:  Secrist, Heather
;  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
;  TITLE OF INVENTION:  AND TREATMENT OF B. MICROTI INFECTION
;  FILE REFERENCE: 210121.426C5
;  CURRENT APPLICATION NUMBER: US/09/569,098A
;  CURRENT FILING DATE:  2000-05-10
;  NUMBER OF SEQ ID NOS: 110
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 45
;   LENGTH: 4223
;   TYPE: DNA
;   ORGANISM: Babesia microti
US-09-569-098A-45

  Query Match             4.3%;  Score 42.8;  DB 4;  Length 4223;
  Best Local Similarity   54.4%;  Pred. No. 0.029;
  Matches   86;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        268 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 327
              ||| ||| ||| |||  ||  || ||| ||| |||  |   || ||  ||| |||  |
Db        656 GGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGTGTGCGTGTGTGCGTGCGTGTG 597

Qy        328 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 387
              || ||  ||| |||  ||  |  ||| ||| |||  |   || ||  ||  ||   |
Db        596 TGCGTGTGTGCGTGCGTGCGTGTGTGCGTGCGTGCGTGTGTGCGTGTGTGTGTGTGTG 537

Qy        388 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 425
              |  ||  ||  ||   |   |  ||  |   || |||
Db        536 TGTGTGTGTGTGTGTGTGTGTGTGTGTTGGGATGGG 499


RESULT 70
US-08-903-800A-4
; Sequence 4, Application US/08903800A
; Patent No. 5935789
;  GENERAL INFORMATION:
;    APPLICANT:  RHEE, Sang-Ki
;    APPLICANT:  CHOI, Eui-Sung
;    APPLICANT:  KIM, Chul-Ho
;    APPLICANT:  SOHN, Jung-Hoon
;    APPLICANT:  KANG, Hyun-Ah
```

```
;  • APPLICANT:  KIM, Hwa-Young
;    TITLE OF INVENTION:  AUTONOMOUSLY REPLICATING SEQUENCES,
;    TITLE OF INVENTION:  GAPDH GENE AND PROMOTER DERIVED FROM
;    TITLE OF INVENTION:  HANSENULA POLYMORPHA, EXPRESSION
;    TITLE OF INVENTION:  VECTORS CONTAINING SAME AND METHOD FOR
;    TITLE OF INVENTION:  THE SELECTION OF TRANSFORMANTS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RHEE, Sang-Ki
;      STREET:  Keukdong Villa Ka-101, Kwangjang-dong,
;      STREET:  Kwangjin-gu
;      CITY:  Seoul
;      STATE:  Seoul
;      COUNTRY:  Republic of Korea
;      ZIP:  143-210
;      ADDRESSEE:  CHOI, Eui-Sung
;      STREET:  Dasol Apt. 102-507, #395-3, Kung-dong,
;      STREET:  Yuseong-gu
;      CITY:  Daejeon
;      STATE:  Daejeon
;      COUNTRY:  Republic of Korea
;      ZIP:  305-335
;      ADDRESSEE:  KIM, Chul-Ho
;      STREET:  Shindonga Apt. 2-1101, Yongjeon-dong, Dong-gu
;      CITY:  Daejeon
;      STATE:  Daejeon
;      COUNTRY:  Republic of Korea
;      ZIP:  300-200
;      ADDRESSEE:  SOHN, Jung-Hoon
;      STREET:  Nuri Apt. 103-506, Wolpyung-dong, Seo-gu
;      CITY:  Daejeon
;      STATE:  Daejeon
;      COUNTRY:  Republic of Korea
;      ZIP:  302-280
;      ADDRESSEE:  KANG, Hyun-Ah
;      STREET:  Kyungseong Keunmaul Apt. 125-1501, Kalma-dong,
;      STREET:  Seo-gu
;      CITY:  Daejeon
;      STATE:  Daejeon
;      COUNTRY:  Republic of Korea
;      ZIP:  302-171
;      ADDRESSEE:  KIM, Hwa-Young
;      STREET:  #3-462, Bukahyun-3-dong, Seodaemun-gu
;      CITY:  Seoul
;      STATE:  Seoul
;      COUNTRY:  Republic of Korea
;      ZIP:  120-190
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5inch, 1.44MB storage
;      COMPUTER:  IBM PC/AT
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/903,800A
;      FILING DATE:  31-JUL-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  KR 97-3173
;      FILING DATE:  31-JAN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:
;      REGISTRATION NUMBER:
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:
;      TELEFAX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  460 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      OTHER INFORMATION:  autonomously replicating sequnce of
;      OTHER INFORMATION:  Hansenula polymorpha DL-1(ATCC 26012)
US-08-903-800A-4

  Query Match             4.2%;  Score 42;  DB 2;  Length 460;
  Best Local Similarity   53.7%;  Pred. No. 0.021;
  Matches   87;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy        217 GGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 276
              || ||| | ||    | | |||  ||  | ||||  | | ||  | | || ||   ||
Db        298 GGATGCGGTCTGAGGATTGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGG 357

Qy        277 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 336
              | | ||  | ||||  | |.||  | | || ||   ||  | | ||  | ||||  | |
Db        358 GGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTG 417

Qy        337 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 378
              ||  | | || ||   ||  | | ||  | ||||  | | ||
Db        418 GCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGC 459


RESULT 71
US-09-252-991A-16275/c
; Sequence 16275, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 16275
;   LENGTH: 465
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16275

  Query Match             4.2%;  Score 42;  DB 4;  Length 465;
  Best Local Similarity   49.4%;  Pred. No. 0.021;
  Matches  163;  Conservative    0;  Mismatches  165;  Indels    2;  Gaps    2;

Qy         82 CTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGA 141
              || ||||  | || |     | |        ||| |   ||| || | |    | || |
Db        397 CTGGCTGCTCGGCGTCGCCGGAGCGTTCGCTGCCGGAGCGCTGGCAGCCGGTGTGGCGGC 338

Qy        142 GGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTG 201
              |||| |||  |||  ||| |  || | | | ||      |     |  | | |  |
Db        337 GGGGGCAGCGGGCGCGCTCGAAGACGCGGCGGGAGCGGCCGGTTTCGCCGCGGCGGGCTT 278

Qy        202 CTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTG 261
                |  | |       || || |||||  | | | | | ||| |  |||  | ||    |
Db        277 GGCTGCCGGCTTCTTCGCCGCAGGCTTTTTCGCA-GCCGGCTTGGCGGCAGGTTTCGCCG 219

Qy        262 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 321
              |  ||| | | ||   |       || ||| ||  |||  ||  | |   | ||  ||||
Db        218 CGGGCTTGGCGGCGGTCGCAGCGGCCGGCTTGGCTGC-GGCAGGTTTCGCAGCAGGCTTC 160

Qy        322 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 381
               ||| | | ||  |||  ||| | | ||| | ||   |||    ||   | |   | |||
Db        159 GCCGCGGCCTTGGCTGCCGGCTTGGTCGCCGGTTTCGCTGCGGCCTTCGCAGTCGGTTTG 100

Qy        382 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
                 || ||| ||  |||  ||| |  ||||
Db         99 GCGGCGGGCTTGGCTGCCGGCTTGGCGGCC 70
```

```
RESULT 72
US-09-298-367B-9/c
; Sequence 9, Application US/09298367B
; Patent No. 6180112
; GENERAL INFORMATION:
;  APPLICANT:  Highlander, Sarah K.
;  APPLICANT:  Federova, Natalie D.
;  TITLE OF INVENTION:  PASTEURELLA HAEMOLYTICA VACCINE
;  FILE REFERENCE:  BCM-03728
;  CURRENT APPLICATION NUMBER: US/09/298,367B
;  CURRENT FILING DATE:  1999-04-22
;  PRIOR APPLICATION NUMBER:  08/834,455
;  PRIOR FILING DATE:  1997-04-15
;  NUMBER OF SEQ ID NOS:  11
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 2445
;   TYPE: DNA
;   ORGANISM: Pasteurella haemolytica
US-09-298-367B-9

  Query Match             4.2%;  Score 42;  DB 3;  Length 2445;
  Best Local Similarity   51.0%;  Pred. No. 0.039;
  Matches   99;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy        488 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 547
              ||||| ||||      |   ||||| ||||      |   ||||| ||||      |
Db        658 TGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 599

Qy        548 TGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 607
              ||||| ||||      |   | ||| ||||      |   ||||| ||||      |
Db        598 TGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGCTG 539

Qy        608 TCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 667
              | ||| ||||      |   ||||| ||||      |   ||||| ||||      |
Db        538 TGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 479

Qy        668 TGCTGGGCTGCTAG 681
              ||||| |||| | |
Db        478 TGCTGTGCTGTTTG 465


RESULT 73
US-09-298-367B-6/c
; Sequence 6, Application US/09298367B
; Patent No. 6180112
; GENERAL INFORMATION:
;  APPLICANT:  Highlander, Sarah K.
;  APPLICANT:  Federova, Natalie D.
;  TITLE OF INVENTION:  PASTEURELLA HAEMOLYTICA VACCINE
;  FILE REFERENCE:  BCM-03728
;  CURRENT APPLICATION NUMBER: US/09/298,367B
;  CURRENT FILING DATE:  1999-04-22
;  PRIOR APPLICATION NUMBER:  08/834,455
;  PRIOR FILING DATE:  1997-04-15
;  NUMBER OF SEQ ID NOS:  11
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 8501
;   TYPE: DNA
;   ORGANISM: Pasteurella haemolytica
US-09-298-367B-6

  Query Match             4.2%;  Score 42;  DB 3;  Length 8501;
  Best Local Similarity   51.0%;  Pred. No. 0.061;
  Matches   99;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy        488 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 547
              ||||| ||||      |   ||||| ||||      |   ||||| ||||      |
Db       2346 TGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 2287

Qy        548 TGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 607
              ||||| ||||      |   | ||| ||||      |   ||||| ||||      |
Db       2286 TGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 2227

Qy        608 TCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 667
              | ||| ||||      |   ||||| ||||      |   ||||| ||||      |
Db       2226 TGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 2167

Qy        668 TGCTGGGCTGCTAG 681
              ||||| |||| | |
Db       2166 TGCTGTGCTGTTTG 2153


RESULT 74
US-09-345-882-1
; Sequence 1, Application US/09345882
; Patent No. 6399373
; GENERAL INFORMATION:
;  APPLICANT: Bougueleret, Lydie
;  TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-
;  TITLE OF INVENTION:  AND POLYMORPHIC MARKERS ASSOCIATED WITH SAID NUCLEIC ACID.
;  FILE REFERENCE: GENSET.031A
;  CURRENT APPLICATION NUMBER: US/09/345,882
;  CURRENT FILING DATE:  1999-06-30
;  PRIOR APPLICATION NUMBER: US 60/091,315
;  PRIOR FILING DATE: 1998-06-30
;  PRIOR APPLICATION NUMBER: US 60/111,909
;  PRIOR FILING DATE: 1998-12-10
;  NUMBER OF SEQ ID NOS: 140
;  SOFTWARE: Patent.pm
; SEQ ID NO 1
;   LENGTH: 162450
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 72794
;   OTHER INFORMATION: 5-124-273  :  polymorphic  base  A  or  G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 88073
;   OTHER INFORMATION: 5-127-261  :  polymorphic  base  A  or  C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 90842
;   OTHER INFORMATION: 99-1437-325  :  polymorphic  base  A  or  G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 93714
;   OTHER INFORMATION: 5-128-60  :  polymorphic  base  deletion  of  GT
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 97122
;   OTHER INFORMATION: 99-1442-224  :  polymorphic  base  G  or  T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 97152
;   OTHER INFORMATION: 5-129-144  :  polymorphic  base  deletion  of  T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 99098
;   OTHER INFORMATION: 5-130-257  :  polymorphic  base  A  or  G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 99117
;   OTHER INFORMATION: 5-130-276  :  polymorphic  base  A  or  G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 103806
;   OTHER INFORMATION: 5-131-395  :  polymorphic  base  A  or  T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 106940
;   OTHER INFORMATION: 5-133-375  :  polymorphic  base  insertion  of  A
;   FEATURE:
```

```
;  NAME/KEY: allele
;  LOCATION: 108106
;  OTHER INFORMATION: 5-135-155  :  polymorphic  base  insertion  of  A
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 108149
;  OTHER INFORMATION: 5-135-198  :  polymorphic  base  insertion  of  GTTT
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 108308
;  OTHER INFORMATION: 5-135-357  :  polymorphic  base  A  or  G
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 108471
;  OTHER INFORMATION: 5-136-174  :  polymorphic  base  C  or  T
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 134134
;  OTHER INFORMATION: 5-140-120  :  polymorphic  base  C  or  T
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 134362
;  OTHER INFORMATION: 5-140-348  :  polymorphic  base  insertion  of  A
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 134374
;  OTHER INFORMATION: 5-140-361  :  polymorphic  base  insertion  of  CA
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 146328
;  OTHER INFORMATION: 5-143-84  :  polymorphic  base  A  or  G
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 146345
;  OTHER INFORMATION: 5-143-101  :  polymorphic  base  A  or  C
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 150329
;  OTHER INFORMATION: 5-145-24  :  polymorphic  base  A  or  G
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 160031
;  OTHER INFORMATION: 5-148-352  :  polymorphic  base  G  or  T
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 72771..72817
;  OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 72771..72817
;  OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 88050..88096
;  OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 88050..88096
;  OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 90819..90865
;  OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 90819..90865
;  OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 93690..93736
;  OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 93690..93736
;  OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 97099..97145
;  OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 97099..97145
;  OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 97130..97177
;  OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 97130..97177
;  OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 99075..99121
;  OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 99075..99121
;  OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 99094..99140
;  OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 99094..99140
;  OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 103783..103828
;  OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 103783..103828
;  OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 106918..106966
;  OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 106918..106966
;  OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 108084..108130
;  OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 108084..108130
;  OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 108127..108177
;  OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 108127..108177
;  OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
;  FEATURE:

  Query Match             4.2%;  Score 42;  DB 4;  Length 162450;
  Best Local Similarity   47.4%;  Pred. No. 0.18;
  Matches  199;  Conservative    0;  Mismatches  210;  Indels   11;  Gaps    2;

Qy        251 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 310
```

```
Db      19045 CTCCTCACTTCCCAGACGGGGTGGTTGCCGGGCGGAGGGGCTCCTCACTTCTCAGATGGG 19104

Qy        311 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 370

Db      19105 GCGGCTGCCGGGCGGAGGGGCTCCTCACTTCTCAGATGGGGCGGCTGCCGGGCGGAGGGG 19164

Qy        371 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 430

Db      19165 CTCCTCACTTCTCAGATGGGGCGGTTGCCAGGTGGAGGGTCTCCCCACTTCTCAGACGGG 19224

Qy        431 AGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGC 490

Db      19225 GCGGCCG-------GGCAGAGACGCTCCTCACCTCCCAGACGGGGTCACGGCCGGGCAGA 19277

Qy        491 TGGGCTGCTAGTA----GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 546

Db      19278 GGCGCTCCTCACATCCCAGACGGGGCGGCGGGGCAGAGGTGCTCCCCACATCTCAGACAA 19337

Qy        547 CTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 606

Db      19338 TGGGCGGCCGGGCAGAGACGCTCCTCACTTCCTAGATGGGATGGCGGCCGGGAAGAGGCG 19397

Qy        607 CTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 666

Db      19398 CTCCTCACTTCCTAGATGGGATGGCGGCCGGGCAGAGACGCTCGTCACTTTCCAGACTGG 19457


RESULT 75
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
;  APPLICANT: Pan, Yang
;  TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
;  TITLE OF INVENTION: AND USES THEREOF
;  FILE REFERENCE: 09404/052001
;  CURRENT APPLICATION NUMBER: US/09/128,155
;  CURRENT FILING DATE: 1998-08-03
;  EARLIER APPLICATION NUMBER: US 60/091,650
;  EARLIER FILING DATE: 1998-07-02
;  EARLIER APPLICATION NUMBER: US 60/054,646
;  EARLIER FILING DATE: 1997-08-04
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 16
;   LENGTH: 152331
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(152331)
;   OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

  Query Match             4.2%;  Score 41.8;  DB 3;  Length 152331;
  Best Local Similarity   53.7%;  Pred. No. 0.2;
  Matches  109;  Conservative    0;  Mismatches   92;  Indels    2;  Gaps    1;

Qy        219 CTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 278

Db     128467 CTGCCACCTTCCTCAGAGCCTGTCTTTCTTCCTTCCTTCCTTCCTTCCTCTTTCTTTTTC 128526

Qy        279 TTGCTCGCTGGCCTGCTTGC--TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 336

Db     128527 TTTCTCTCTCTCTTTCTTTCTTTTCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT 128586

Qy        337 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 396

Db     128587 TCTTTCTTTCTTTCTTTCTTCCTTTCTTTCTCTTTCTCTCTTTCTCTCTTTCTCTTTCTT 128646

Qy        397 GCTTGCTCGCTGGCCTGCTTGCT 419

Db     128647 TCTTTCTCTCTCCCTCCCTCCCT 128669


RESULT 76
US-09-734-674-3/c
; Sequence 3, Application US/09734674
; Patent No. 6498022
; GENERAL INFORMATION:
;  APPLICANT: WEI, Ming-Hui et al
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL001018
;  CURRENT APPLICATION NUMBER: US/09/734,674
;  CURRENT FILING DATE:  2000-12-13
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 202001
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(202001)
;   OTHER INFORMATION: n = A,T,C or G
US-09-734-674-3

  Query Match             4.2%;  Score 41.6;  DB 4;  Length 202001;
  Best Local Similarity   53.8%;  Pred. No. 0.25;
  Matches   86;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        255 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 314

Db      73047 TTCTTTCCTTCCTTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC 72988

Qy        315 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 374

Db      72987 TTTCTTTCTTTCTTTCTCTCTCTCTCTCTCTCTTTCTTTCTTTCTTTTTCTTCCTTTTCT 72928

Qy        375 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 414

Db      72927 TTCTTTTCTTTCTTTCACTCTGTCACCCACACTGGAGTGC 72888


RESULT 77
US-09-803-671B-3
; Sequence 3, Application US/09803671B
; Patent No. 6582946
; GENERAL INFORMATION:
;  APPLICANT: WEBSTER, Marion et al
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001161
;  CURRENT APPLICATION NUMBER: US/09/803,671B
;  CURRENT FILING DATE:  2001-03-12
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 64467
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(64467)
;   OTHER INFORMATION: n = A,T,C or G
US-09-803-671B-3

  Query Match             4.1%;  Score 41.4;  DB 4;  Length 64467;
  Best Local Similarity   57.3%;  Pred. No. 0.19;
  Matches   75;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
```

```
Db        249 TTCTTTCTTTCTCTCTCTCTTTCTTTCTTTCTCTCTTTCTTTCTTTCTTTCTTTATTTCT 308

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371

Db        309 CTGTCTCTTTCTTTTCCCTCTCCTTCCTTCCTTCCTTCCTTTCTCTCTCTCTCTTTCT 368

Qy        372 TGCTTGCTTGC 382

Db        369 TTCTTCCTTTC 379


RESULT 78
US-08-903-800A-1
; Sequence 1, Application US/08903800A
; Patent No. 5935789
;  GENERAL INFORMATION:
;    APPLICANT:  RHEE, Sang-Ki
;    APPLICANT:  CHOI, Eui-Sung
;    APPLICANT:  KIM, Chul-Ho
;    APPLICANT:  SOHN, Jung-Hoon
;    APPLICANT:  KANG, Hyun-Ah
;    APPLICANT:  KIM, Hwa-Young
;    TITLE OF INVENTION:  AUTONOMOUSLY REPLICATING SEQUENCES,
;    TITLE OF INVENTION:  GAPDH GENE AND PROMOTER DERIVED FROM
;    TITLE OF INVENTION:  HANSENULA POLYMORPHA, EXPRESSION
;    TITLE OF INVENTION:  VECTORS CONTAINING SAME AND METHOD FOR
;    TITLE OF INVENTION:  THE SELECTION OF TRANSFORMANTS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RHEE, Sang-Ki
;      STREET:  Keukdong Villa Ka-101, Kwangjang-dong,
;      STREET:  Kwangjin-gu
;      CITY:  Seoul
;      STATE:  Seoul
;      COUNTRY:  Republic of Korea
;      ZIP:  143-210
;      ADDRESSEE:  CHOI, Eui-Sung
;      STREET:  Dasol Apt. 102-507, #395-3, Kung-dong,
;      STREET:  Yuseong-gu
;      CITY:  Daejeon
;      STATE:  Daejeon
;      COUNTRY:  Republic of Korea
;      ZIP:  305-335
;      ADDRESSEE:  KIM, Chul-Ho
;      STREET:  Shindonga Apt. 2-1101, Yongjeon-dong, Dong-gu
;      CITY:  Daejeon
;      STATE:  Daejeon
;      COUNTRY:  Republic of Korea
;      ZIP:  300-200
;      ADDRESSEE:  SOHN, Jung-Hoon
;      STREET:  Nuri Apt. 103-506, Wolpyung-dong, Seo-gu
;      CITY:  Daejeon
;      STATE:  Daejeon
;      COUNTRY:  Republic of Korea
;      ZIP:  302-280
;      ADDRESSEE:  KANG, Hyun-Ah
;      STREET:  Kyungseong Keunmaul Apt. 125-1501, Kalma-dong,
;      STREET:  Seo-gu
;      CITY:  Daejeon
;      STATE:  Daejeon
;      COUNTRY:  Republic of Korea
;      ZIP:  302-171
;      ADDRESSEE:  KIM, Hwa-Young
;      STREET:  #3-462, Bukahyun-3-dong, Seodaemun-gu
;      CITY:  Seoul
;      STATE:  Seoul
;      COUNTRY:  Republic of Korea
;      ZIP:  120-190
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5inch, 1.44MB storage
;      COMPUTER:  IBM PC/AT
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/903,800A
;      FILING DATE:  31-JUL-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  KR 97-3173
;      FILING DATE:  31-JAN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:
;      REGISTRATION NUMBER:
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:
;      TELEFAX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1227 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      OTHER INFORMATION:  autonomously replicating sequnce of
;      OTHER INFORMATION:  Hansenula polymorpha DL-1(ATCC 26012)
US-08-903-800A-1

  Query Match             4.1%;  Score 41;  DB 2;  Length 1227;
  Best Local Similarity   50.2%;  Pred. No. 0.056;
  Matches  101;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy        189 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 248

Db        501 GATGCGGTGAGGGGAGCGGACGCGGTTGGTTTTAGGATGCGGTCTGAGGGTGGCGGGGTG 560

Qy        249 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 308

Db        561 GCGGGGTGGCGGGGTGGTGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGG 620

Qy        309 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 368

Db        621 GGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTG 680

Qy        369 GCCTGCTTGCTTGCTCGCTGG 389

Db        681 GCGGGGTGGCGTGATCATCGG 701


RESULT 79
US-07-922-723A-7/c
; Sequence 7, Application US/07922723A
; Patent No. 5369004
;  GENERAL INFORMATION:
;    APPLICANT:  Drs. Mihael H. Polymeropoulos
;    APPLICANT:  and Carl R. Merril
;    TITLE OF INVENTION:  FIVE HIGHLY INFORMATIVE
;    TITLE OF INVENTION:  REPEAT POLYMORPHIC DNA MARKERS
;    NUMBER OF SEQUENCES:  73
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lowe, Price, LeBlanc & Becker
;      STREET:  Suite 300, 99 Canal Center Plaza
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  DOS Text File
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/922,723A
;      FILING DATE:
```

```
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  D.J. Mills
;      REGISTRATION NUMBER:  34506
;      REFERENCE/DOCKET NUMBER:  717081B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703 684 1111
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  291
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-07-922-723A-7

  Query Match             4.1%;  Score 40.8;  DB 1;  Length 291;
  Best Local Similarity   53.8%;  Pred. No. 0.038;
  Matches   84;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||   ||  |
Db        188 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTCTTT 129

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
                ||| ||| ||   |       || ||| ||  ||  | | ||  ||  ||| ||  |
Db        128 CTCTTTCTTTCTTTTTCTTTCTTTTTCTTCCTTCCTTCCTTTCTCTCTCTCTCTCTTTCT 69

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 407
              | ||| ||  ||| ||      ||| ||| ||  ||
Db         68 TTCTTTCTAACTCTCTTTGTCTCTTTCTTTCTTTCT 33


RESULT 80
US-07-799-828C-7/c
; Sequence 7, Application US/07799828C
; Patent No. 5378602
;  GENERAL INFORMATION:
;    APPLICANT:  Drs. Carl R. Merril and
;    APPLICANT:  Mihael H. Polymeropoulos
;    TITLE OF INVENTION:  TWENTY SEVEN HIGHLY INFORMATIVE
;    TITLE OF INVENTION:  MICROSATELLITE REPEAT
;    TITLE OF INVENTION:  POLYMORPHIC DNA MARKERS
;    NUMBER OF SEQUENCES:  63
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lowe, Price, LeBlanc & Becker
;      STREET:  Suite 300, 99 Canal Center Plaza
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  DOS Text File
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/799,828C
;      FILING DATE:  19911127
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  D.J. Mills
;      REGISTRATION NUMBER:  34,506
;      REFERENCE/DOCKET NUMBER:  717081A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703 684 1111
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  291
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-07-799-828C-7

  Query Match             4.1%;  Score 40.8;  DB 1;  Length 291;
  Best Local Similarity   53.8%;  Pred. No. 0.038;
  Matches   84;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||   ||  |
Db        188 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTCTTT 129

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
                ||| ||| ||   |       || ||| ||  ||  | | ||  ||  ||| ||  |
Db        128 CTCTTTCTTTCTTTTTCTTTCTTTTTCTTCCTTCCTTCCTTTCTCTCTCTCTCTCTTTCT 69

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 407
              | ||| ||  ||| ||      ||| ||| ||  ||
Db         68 TTCTTTCTAACTCTCTTTGTCTCTTTCTTTCTTTCT 33


RESULT 81
US-08-074-275-7/c
; Sequence 7, Application US/08074275
; Patent No. 5468610
;  GENERAL INFORMATION:
;    APPLICANT:  Drs. Carl R. Merril and
;    APPLICANT:  Mihael H. Polymeropoulos
;    TITLE OF INVENTION:  THREE HIGHLY INFORMATIVE REPEAT
;    TITLE OF INVENTION:  POLYMORPHIC DNA MARKERS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lowe, Price, LeBlanc & Becker
;      STREET:  Suite 300, 99 Canal Center Plaza
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  DOS Text File
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/074,275
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/707,501
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  J.G. Mullins
;      REGISTRATION NUMBER:  33073
;      REFERENCE/DOCKET NUMBER:  717081
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703 684 1111
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  291
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-074-275-7

  Query Match             4.1%;  Score 40.8;  DB 1;  Length 291;
  Best Local Similarity   53.8%;  Pred. No. 0.038;
  Matches   84;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||   ||  |
Db        188 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTCTTT 129

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
                ||| ||| ||   |       || ||| ||  ||  | | ||  ||  ||| ||  |
```

```
Db        128 CTCTTTCTTTCTTTTTCTTTCTTTTTCTTCCTTCCTTCCTTTCTCTCTCTCTCTTTCT 69

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 407
              | ||| ||  ||| ||      ||| ||| ||  ||
Db         68 TTCTTTCTAACTCTCTTTGTCTCTTTCTTTCTTTCT 33


RESULT 82
US-08-480-366-7/c
; Sequence 7, Application US/08480366
; Patent No. 5721100
;  GENERAL INFORMATION:
;    APPLICANT:  Drs. Carl R. Merril and
;    APPLICANT:  Mihael H. Polymeropoulos
;    TITLE OF INVENTION:  THREE HIGHLY INFORMATIVE REPEAT
;    TITLE OF INVENTION:  POLYMORPHIC DNA MARKERS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lowe, Price, LeBlanc & Becker
;      STREET:  Suite 300, 99 Canal Center Plaza
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  DOS Text File
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,366
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  J.G. Mullins
;      REGISTRATION NUMBER:  33073
;      REFERENCE/DOCKET NUMBER:  717081
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703 684 1111
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  291
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-480-366-7

  Query Match             4.1%;  Score 40.8;  DB 1;  Length 291;
  Best Local Similarity   53.8%;  Pred. No. 0.038;
  Matches   84;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
              | ||| ||| ||  || ' | | ||| ||| ||  ||  | | ||| ||   ||  |
Db        188 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTCTTT 129

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
                ||| ||| ||   |       || ||| ||  ||  | | ||  ||  ||| ||  |
Db        128 CTCTTTCTTTCTTTTTCTTTCTTTTTCTTCCTTCCTTCCTTTCTCTCTCTCTCTCTTTCT 69

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 407
              | ||| ||  ||| ||      ||| ||| ||  ||
Db         68 TTCTTTCTAACTCTCTTTGTCTCTTTCTTTCTTTCT 33


RESULT 83
US-07-952-277A-7/c
; Sequence 7, Application US/07952277A
; Patent No. 5861504
;  GENERAL INFORMATION:
;    APPLICANT:  Drs. Mihael H. Polymeropoulos
;    APPLICANT:  and Carl R. Merril
;    TITLE OF INVENTION:  ELEVEN HIGHLY INFORMATIVE
;    TITLE OF INVENTION:  REPEAT POLYMORPHIC DNA MARKERS
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lowe, Price, LeBlanc & Becker
;      STREET:  Suite 300, 99 Canal Center Plaza
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  DOS Text File
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/952,277A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  D.J. Mills
;      REGISTRATION NUMBER:  34506
;      REFERENCE/DOCKET NUMBER:  717081C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703 684 1111
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  291
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-07-952-277A-7

  Query Match             4.1%;  Score 40.8;  DB 2;  Length 291;
  Best Local Similarity   53.8%;  Pred. No. 0.038;
  Matches   84;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||   ||  |
Db        188 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTCTTT 129

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
                ||| ||| ||   |       || ||| ||  ||  | | ||  ||  ||| ||  |
Db        128 CTCTTTCTTTCTTTTTCTTTCTTTTTCTTCCTTCCTTCCTTTCTCTCTCTCTCTCTTTCT 69

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 407
              | ||| ||  ||| ||      ||| ||| ||  ||
Db         68 TTCTTTCTAACTCTCTTTGTCTCTTTCTTTCTTTCT 33


RESULT 84
US-08-903-800A-2
; Sequence 2, Application US/08903800A
; Patent No. 5935789
;  GENERAL INFORMATION:
;    APPLICANT:  RHEE, Sang-Ki
;    APPLICANT:  CHOI, Eui-Sung
;    APPLICANT:  KIM, Chul-Ho
;    APPLICANT:  SOHN, Jung-Hoon
;    APPLICANT:  KANG, Hyun-Ah
;    APPLICANT:  KIM, Hwa-Young
;    TITLE OF INVENTION:  AUTONOMOUSLY REPLICATING SEQUENCES,
;    TITLE OF INVENTION:  GAPDH GENE AND PROMOTER DERIVED FROM
;    TITLE OF INVENTION:  HANSENULA POLYMORPHA, EXPRESSION
;    TITLE OF INVENTION:  VECTORS CONTAINING SAME AND METHOD FOR
;    TITLE OF INVENTION:  THE SELECTION OF TRANSFORMANTS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RHEE, Sang-Ki
;      STREET:  Keukdong Villa Ka-101, Kwangjang-dong,
;      STREET:  Kwangjin-gu
;      CITY:  Seoul
;      STATE:  Seoul
```

```
;       COUNTRY:  Republic of Korea
;       ZIP:  143-210
;       ADDRESSEE:  CHOI, Eui-Sung
;       STREET:  Dasol Apt. 102-507, #395-3, Kung-dong,
;       STREET:  Yuseong-gu
;       CITY:  Daejeon
;       STATE:  Daejeon
;       COUNTRY:  Republic of Korea
;       ZIP:  305-335
;       ADDRESSEE:  KIM, Chul-Ho
;       STREET:  Shindonga Apt. 2-1101, Yongjeon-dong, Dong-gu
;       CITY:  Daejeon
;       STATE:  Daejeon
;       COUNTRY:  Republic of Korea
;       ZIP:  300-200
;       ADDRESSEE:  SOHN, Jung-Hoon
;       STREET:  Nuri Apt. 103-506, Wolpyung-dong, Seo-gu
;       CITY:  Daejeon
;       STATE:  Daejeon
;       COUNTRY:  Republic of Korea
;       ZIP:  302-280
;       ADDRESSEE:  KANG, Hyun-Ah
;       STREET:  Kyungseong Keunmaul Apt. 125-1501, Kalma-dong,
;       STREET:  Seo-gu
;       CITY:  Daejeon
;       STATE:  Daejeon
;       COUNTRY:  Republic of Korea
;       ZIP:  302-171
;       ADDRESSEE:  KIM, Hwa-Young
;       STREET:  #3-462, Bukahyun-3-dong, Seodaemun-gu
;       CITY:  Seoul
;       STATE:  Seoul
;       COUNTRY:  Republic of Korea
;       ZIP:  120-190
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.5inch, 1.44MB storage
;       COMPUTER:  IBM PC/AT
;       OPERATING SYSTEM:  MS-DOS
;       SOFTWARE:  Word Perfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/903,800A
;       FILING DATE:  31-JUL-1997
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  KR 97-3173
;       FILING DATE:  31-JAN-1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME:
;       REGISTRATION NUMBER:
;       REFERENCE/DOCKET NUMBER:
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:
;       TELEFAX:
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  453 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  genomic DNA
;     FEATURE:
;       OTHER INFORMATION:  autonomously replicating sequnce of
;       OTHER INFORMATION:  Hansenula polymorpha DL-1(ATCC 26012)
US-08-903-800A-2

  Query Match           4.0%;  Score 40.4;  DB 2;  Length 453;
  Best Local Similarity   48.3%;  Pred. No. 0.057;
  Matches  113;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy        468 TGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 527
              | | ||   || | |||    | | || ||  | |  | ||  ||| |||    ||||
Db        217 TAAAGGTGAGAAATTAGATGGGAGGAGCGGCAGGAAACGGTGTAGGGATGCGGTGAGGGG 276

Qy        528 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGC 587
              || |    | | |       | || | |   ||   |||     ||| |   |  |||
Db        277 AGCGGACGCGGTTGGTTTTAGGATGCGGTCTGAGGGTGGCGGGGTGGCGGGGTGGCGGGG 336

Qy        588 TGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 647
              || |||| ||   |  |||     ||| |   |  ||| ||  ||| ||   |   |||
Db        337 TGGTGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGC 396

Qy        648 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTG 701
              |  ||| |   |  ||| ||  ||| ||   |  ||| |  ||  ||  ||
Db        397 GGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTG 450


RESULT 85
US-08-365-486A-11
; Sequence 11, Application US/08365486A
; Patent No. 5834306
;  GENERAL INFORMATION:
;    APPLICANT:  Webster, Keith A.
;    APPLICANT:  Bishopric, Nanette H.
;    TITLE OF INVENTION:  Tissue Specific Hypoxia Regulated
;    TITLE OF INVENTION:  Therapeutic Constructs
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/365,486A
;      FILING DATE:  23-DEC-1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sholtz, Charles K.
;      REGISTRATION NUMBER:  38,615
;      REFERENCE/DOCKET NUMBER:  8255-0018
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1679 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  Mouse alpha MHC promoter fragment
US-08-365-486A-11

  Query Match           4.0%;  Score 40.4;  DB 2;  Length 1679;
  Best Local Similarity   56.9%;  Pred. No. 0.092;
  Matches   74;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy        239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
              || | |  |  |   | | ||  | |   | | ||| ||| ||  ||  | | ||| |
Db        976 CTTCACCTTTTCTCCCCTTCTCATTTGTATTCATCCTTCCTTCCTTCCTTCCTTCCTTCC 1035

Qy        299 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 358
              || ||  ||  | | ||| ||| ||| ||  | | ||| ||| ||  ||  | | ||| |
Db       1036 TTCCTTCCTTCCTTCCTTCCTTTCTCCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCC 1095

Qy        359 TTGCTCGCTG 368
              || ||  |||
```

```
Db *      1096 TTCCTTCCTG 1105


RESULT 86
US-08-880-342-11
; Sequence 11, Application US/08880342
; Patent No. 6218179
;  GENERAL INFORMATION:
;    APPLICANT:  Webster, Keith A.
;    APPLICANT:  Bishopric, Nanette H.
;    APPLICANT:  Murphy, Brian
;    APPLICANT:  Laderoute, Keith R.
;    APPLICANT:  Green, Christopher J.
;    TITLE OF INVENTION:  Tissue Specific Hypoxia Regulated
;    TITLE OF INVENTION:  Therapeutic Constructs
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/880,342
;      FILING DATE:  23-JUN-1997
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/IB95/00996
;      FILING DATE:  13-NOV-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/365,486
;      FILING DATE:  23-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sholtz, Charles K.
;      REGISTRATION NUMBER:  38,615
;      REFERENCE/DOCKET NUMBER:  8255-0018.30
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1679 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  Mouse alpha MHC promoter fragment
US-08-880-342-11

  Query Match             4.0%;  Score 40.4;  DB 3;  Length 1679;
  Best Local Similarity   56.9%;  Pred. No. 0.092;
  Matches   74;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy        239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
              || | |  |  |   | | ||   | |    | | ||| ||| ||  ||  | | ||| |
Db        976 CTTCACCTTTTCTCCCCTTCTCATTTGTATTCATCCTTCCTTCCTTCCTTCCTTCCTTCC 1035

Qy        299 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 358
              || ||  ||  | | ||| ||| ||| ||  | | ||| ||| ||  ||  | | ||| |
Db       1036 TTCCTTCCTTCCTTCCTTCCTTTCTCCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCC 1095

Qy        359 TTGCTCGCTG 368
              || ||  |||
Db       1096 TTCCTTCCTG 1105


RESULT 87
US-09-146-053-5/c
; Sequence 5, Application US/09146053A
; Patent No. 6399349
; GENERAL INFORMATION:
;  APPLICANT: Ryan, James W.
;  APPLICANT: Sprinkle, Terry Joe Curtis
;  APPLICANT: Venema, Richard C.
;  TITLE OF INVENTION: Human Aminopeptidase P Gene
;  FILE REFERENCE: MCG103
;  CURRENT APPLICATION NUMBER: US/09/146,053A
;  CURRENT FILING DATE: 1998-09-02
;  EARLIER APPLICATION NUMBER: 60/057,854
;  EARLIER FILING DATE: 1997-09-02
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 44453
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-146-053-5

  Query Match             4.0%;  Score 40;  DB 4;  Length 44453;
  Best Local Similarity   50.8%;  Pred. No. 0.39;
  Matches  120;  Conservative    0;  Mismatches  115;  Indels    1;  Gaps    1;

Qy        243 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 302
              || ||  | | ||| ||| ||  ||  |   ||| ||| ||  ||  | | ||| ||| |
Db      20968 TCTCTTTCTTTCTTTCTTTCTTTCTTTCTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC 20909

Qy        303 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 362
              |  ||  | | ||| ||| |||  |    | ||| | |  || ||  |   ||  |   |
Db      20908 TTTCTTTCTTTCTTTCTTTCTCTTTCCTTTCCTTTCAT-TTCCCTCCCTCCCTCCCCCCC 20850

Qy        363 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAG 422
              | ||  | | ||| ||| ||  ||  | | ||| |  | | |     |    || ||
Db      20849 ACTCTCTCTTCCTTCCTTCCTTTCTTTCTTTCTTCCAAGATGGGGTCTCACTATGTTGCC 20790

Qy        423 GGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGA 478
              | | | || |  |      | |  ||| |     ||  ||  ||   | |||||
Db      20789 CAGGCTGGAGTGAAGTGGCTGTTCACAGGTGTGATCACAGCTCACTATGGCCTTGA 20734


RESULT 88
US-08-458-434A-6/c
; Sequence 6, Application US/08458434A
; Patent No. 6083690
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Stephen E.
;    APPLICANT:  Mundy M.D., Gregory R.
;    APPLICANT:  Gosh-Choudhury Ph.D., Nandini
;    APPLICANT:  Feng Ph.D., Jian Q.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR IDENTIFYING
;    TITLE OF INVENTION:  OSTEOGENIC AGENTS
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  James C. Weseman, Esq.
;      STREET:  401 B. Street, Suite 1700
;      CITY:  San Diego
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,434A
;      FILING DATE:
;      CLASSIFICATION:  435
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weseman, James C.
;      REGISTRATION NUMBER:  30,507
;      REFERENCE/DOCKET NUMBER:  P00060US0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 699-3604
;      TELEFAX:  619-236-1048
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15144 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-458-434A-6

  Query Match             3.9%;  Score 39.2;  DB 3;  Length 15144;
  Best Local Similarity   55.0%;  Pred. No. 0.43;
  Matches   77;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        243 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 302
              ||  |  | | ||| ||| ||        |   |  ||| | |  |  | | | | ||  |
Db       7966 TCCTTTTCTTCCTTCCTTTCTTTTCTTTCCTTTCCCTTTCCCTTTCCCTTCCCTTCTCTC 7907

Qy        303 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 362
              || ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| |
Db       7906 TCTCTCCCTTCCTTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC 7847

Qy        363 TCGCTGGCCTGCTTGCTTGC 382
              |  ||  | | ||| ||| |
Db       7846 TTTCTTCCTTCCTTCCTTTC 7827


RESULT 89
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             3.9%;  Score 39;  DB 3;  Length 4403765;
  Best Local Similarity   45.3%;  Pred. No. 3.8;
  Matches  179;  Conservative    0;  Mismatches  215;  Indels    1;  Gaps    1;

Qy        218 GCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 277
              || |  |||  ||      | | || || |||||| | ||||    | | || |    |
Db    1121524 GCCGACGGCCCGCGGAACAGCCAGCACGGTGGCCTCCCTGCTGCTGCCCGGGGCGCAATC 1121465

Qy        278 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 337
              |   || |  ||||  |||     |||| ||  | |||  |       ||| ||  |
Db    1121464 GGCGGCGGCCGCTCTGCCGGCTGTGGGGCTGCCCGCCCTGCCGGGCGTGACGCCGGCGAG 1121405

Qy        338 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 397
              |||||||      |   ||   |  ||| |  || ||  |  || |  | ||||  |

Db    1121404 TCTGCTCGCGATGGCGGCGGCCGCCGGCCTGCCAGGCGTGGGTTTTC-CGAGTCTGCCAG 1121346

Qy        398 CTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGG 457
              |   |||| | ||  | || ||  |      |   |    |   ||  | | | |   |
Db    1121345 GTGTGTCGCCGACCGACCTGATGGCGATGGCGGCCGCGGCCGGTCTGCCGACGAGTTTGC 1121286

Qy        458 TACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTG 517
              | ||| | |     | |   |  | ||||  ||       |   ||  |   | |
Db    1121285 CGGGTCTGGCGGGGATGTCGCCGGCGGAGCTGACGGCGCTGGTGGCAGGTGGTCTGCCGA 1121226

Qy        518 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTG 577
                  ||| ||||| |||| |  || ||| |||     |   ||  |  |||||
Db    1121225 TGCTGGCGGCCGCTGGTCTGCCGGCTGGTCTGGCGGGGGTGGATCCGGCGACGCTGGCTG 1121166

Qy        578 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTG 612
              |      | | ||   |  |  ||  || || | |
Db    1121165 CGGCGCTGCCGGCGCTGGCGGCAGGTGGTCTACCG 1121131


RESULT 90
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             3.9%;  Score 39;  DB 3;  Length 4411529;
  Best Local Similarity   45.3%;  Pred. No. 3.8;
  Matches  179;  Conservative    0;  Mismatches  215;  Indels    1;  Gaps    1;

Qy        218 GCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 277
              || |  |||  ||      | | || || |||||| | ||||    | | || |    |
Db    1121502 GCCGACGGCCCGCGGAACAGCCAGCACGGTGGCCTCCCTGCTGCTGCCCGGGGCGCAATC 1121443

Qy        278 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 337
              |   || |  ||||  |||     |||| ||  | |||  |       ||| ||  |
Db    1121442 GGCGGCGGCCGCTCTGCCGGCTGTGGGGCTGCCCGCCCTGCCGGGCGTGACGCCGGCGAG 1121383

Qy        338 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 397
              |||||||      |   ||   |  ||| |  || ||  |  || |  | ||||  |

Db    1121382 TCTGCTCGCGATGGCGGCGGCCGCCGGCCTGCCAGGCGTGGGTTTTC-CGAGTCTGCCAG 1121324

Qy        398 CTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGG 457
              |   |||| | ||  | || ||  |      |   |    |   ||  | | | |   |
Db    1121323 GTGTGTCGCCGACCGACCTGATGGCGATGGCGGCCGCGGCCGGTCTGCCGACGAGTTTGC 1121264

Qy        458 TACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTG 517
              | ||| | |     | |   |  | ||||  ||       |   ||  |   | |
Db    1121263 CGGGTCTGGCGGGGATGTCGCCGGCGGAGCTGACGGCGCTGGTGGCAGGTGGTCTGCCGA 1121204

Qy        518 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTG 577
                  ||| ||||| |||| |  || ||| |||     |   ||  |  |||||
Db    1121203 TGCTGGCGGCCGCTGGTCTGCCGGCTGGTCTGGCGGGGGTGGATCCGGCGACGCTGGCTG 1121144

Qy        578 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTG 612
              |      | | ||   |  |  ||  || || | |
Db    1121143 CGGCGCTGCCGGCGCTGGCGGCAGGTGGTCTACCG 1121109
```

```
RESULT 91
US-08-332-766A-4
; Sequence 4, Application US/08332766A
; Patent No. 5843647
;  GENERAL INFORMATION:
;    APPLICANT:  JEFFREYS, Alec J.
;    APPLICANT:  ARMOUR, John
;    TITLE OF INVENTION:  SIMPLE TANDEM REPEATS
;    NUMBER OF SEQUENCES:  125
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D. C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/332,766A
;      FILING DATE:  01-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9326052.9
;      FILING DATE:  21-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BIRD, Donald J.
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  217211/M94/0434/GB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  217 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-332-766A-4

  Query Match             3.9%;  Score 38.8;  DB 2;  Length 217;
  Best Local Similarity   52.5%;  Pred. No. 0.12;
  Matches   85;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy        242 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 301
              ||| ||  | | ||| ||| ||| ||  |   ||   |  ||| ||  | | ||  |
Db         53 CTCTCTTTCTTCCTTCCTTCCTCCCTCCCTCCCTCCTTCCCTCCCTCCCTTCCTCCCGCC 112

Qy        302 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 361
              |||  |  | |      |  | |     | | ||| ||   ||  |  | | ||| |||
Db        113 CTCCTTCCCTTTTCCCTTCCCCCTTCCTCTTTCTTTCTCTTTCTTTCTCTTTCTTTCTTA 172

Qy        362 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403
              ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||
Db        173 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 214


RESULT 92
US-08-623-906A-6/c
; Sequence 6, Application US/08623906A
; Patent No. 5874217
;  GENERAL INFORMATION:
;    APPLICANT:  Stevenson, Tamara
;    APPLICANT:  Dvorak, Jan
;    APPLICANT:  Halverson, Joy
;    TITLE OF INVENTION:  Microsatellite Sequences for Canine
;    TITLE OF INVENTION:  Genotyping
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/623,906A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sherwood, Pamela J.
;      REGISTRATION NUMBER:  36,677
;      REFERENCE/DOCKET NUMBER:  A-62282/BIR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  454 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..133
;      OTHER INFORMATION:  /note= "Nucleotides 1-133 are
;      OTHER INFORMATION:  unique flanking sequence"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  134..307
;      OTHER INFORMATION:  /note= "Nucleotides 134-207 are
;      OTHER INFORMATION:  repeat sequence".
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  308..454
;      OTHER INFORMATION:  /note= "Nucleotides 308-454 are
;      OTHER INFORMATION:  unique flanking sequence"
US-08-623-906A-6

  Query Match             3.9%;  Score 38.8;  DB 2;  Length 454;
  Best Local Similarity   56.0%;  Pred. No. 0.15;
  Matches   94;  Conservative    0;  Mismatches   72;  Indels    2;  Gaps    1;

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db        305 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 246

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
              | ||| |||  |  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db        245 TTCTTTCTT--TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 188

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              | ||| |||  ||  |    |  ||  ||  | | |    |  || ||
Db        187 TTCTTTCTTTTTCTTTTCTTTTTTTCTTTTTTTTGGTCTTTTTTTTTCT 140


RESULT 93
US-09-690-454-21/c
; Sequence 21, Application US/09690454
; Patent No. 6531447
; GENERAL INFORMATION:
;  APPLICANT: Steven M. Ruben, et al.
;  TITLE OF INVENTION: 32 Human Secreted Proteins
```

```
;  FILE REFERENCE: PZ006P1
;  CURRENT APPLICATION NUMBER: US/09/690,454
;  CURRENT FILING DATE:  2000-10-18
;  PRIOR APPLICATION NUMBER: 09/189,144
;  PRIOR FILING DATE: 1998-11-10
;  PRIOR APPLICATION NUMBER: 60/044,039
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/048,093
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/048,190
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/050,935
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/048,101
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/048,356
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/056,250
;  PRIOR FILING DATE: August 29, 1997
;  PRIOR APPLICATION NUMBER: 60/056,296
;  PRIOR FILING DATE: August 29, 1997
;  PRIOR APPLICATION NUMBER: 60/056,293
;  PRIOR FILING DATE: August 29, 1997
;  NUMBER OF SEQ ID NOS: 229
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
;   LENGTH: 1431
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-690-454-21

  Query Match             3.9%;  Score 38.8;  DB 4;  Length 1431;
  Best Local Similarity   51.1%;  Pred. No. 0.23;
  Matches   91;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy        242 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 301
              | | || |    ||| |  | || |||   |  |  |||| |||   || . | |  ||
Db       1371 CGCTCTCGTTCCCTTTCACGATCTCTGCTTTTTTCTCTTGATCGAATGCGATCCTTATTA 1312

Qy        302 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 361
              |   |||  || ||| ||   || | | |  | |  ||  |||| |  |   || | |
Db       1311 CGATCTGAGCTCCTTTCTCTATCCCGGTCACGATCTCTCTCTCGATCCCGATCTCGGTCT 1252

Qy        362 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              ||| ||    | |   |  | || ||  || | |  | | | ||||  | | ||| ||
Db       1251 CTCTCTTTTGTCCGGTCACGGTCCCTATCCCGTTCTCGTTCCCGCTCCCGTTCTTTCT 1194


RESULT 94
US-09-741-154-3/c
; Sequence 3, Application US/09741154
; Patent No. 6437110
; GENERAL INFORMATION:
;  APPLICANT: BEASLEY, Ellen M. et al
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001061
;  CURRENT APPLICATION NUMBER: US/09/741,154
;  CURRENT FILING DATE:  2000-12-21
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 16389
;   TYPE: DNA
;   ORGANISM: Human
US-09-741-154-3

  Query Match             3.9%;  Score 38.6;  DB 4;  Length 16389;
  Best Local Similarity   46.0%;  Pred. No. 0.65;
  Matches  238;  Conservative    0;  Mismatches  274;  Indels    5;  Gaps    3;

Qy        207 GGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGC 266
              | ||  | | ||  | ||| | | | |  ||   | | | ||    | | | | | ||
Db      16149 GTGAATGGATGGTGGGTGGGTAGGTGGATGGAGGGATGGATGAATGGGTGGGTAGATCAA 16090

Qy        267 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 326
              |||     | | | |   | |||     | |  ||   |   |  ||  ||  ||   |
Db      16089 TGGAACAGTGGATGGAGGGATGGATGAATAGAGTGGATGGGTGGGTGAATGAATGGATGG 16030

Qy        327 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 386
              | ||  ||       |   |  ||   | | | | | |||   | | | | | |
Db      16029 GTACATGGATGGACATGGGAATGGGTGGGGGATGGATGGGTGGGGAGATGGATAGATAAG 15970

Qy        387 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT--GAGGGGACAGTAGGGCTGCTTGCTT 444
              |||   | ||| | | | | ||      | | |  | ||||||||  ||   | | | |
Db      15969 TGGATGGGTTGGTGGATGGATGCATGCATGGATGGGTGGGGACAGGGGGATGGGTGGGTG 15910

Qy        445 GCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAG 504
              | |    ||| || | |   |   ||||| | |||   |||   | ||||    | | |
Db      15909 GGTGGATGGATGGAAGGATGGATGGATGGATGGATGAAAGGATGGATGGGTATGTGGATG 15850

Qy        505 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 564
              |      ||      | | ||    | |   ||  |||  || || | ||  | | |  |
Db      15849 GATGATGGGTAGATAAATGGGTGGACAGATGTGTGAGT--GGATGGTAGGTGGATGGATG 15792

Qy        565 GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 624
              |     ||| | | | | | || |||  || ||   | | || |    || ||   |
Db      15791 GATGGGTGGACAGGT-GGATGGATGCATGGGTGGACGGATGGATGGATGGATGGATGATT 15733

Qy        625 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 684
              || ||   || ||    |   || || |||| | | | | |    |||   |   | |
Db      15732 GGGTGGATGGATGGATGAGTGGATGGTGGGTAGGTGGATGAGTGCATGGTGGGTGCGATG 15673

Qy        685 GGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGG 721
              |   |   | ||| | | | | |   | | |||||||
Db      15672 GATAGGTGGGTGGGTGGATGGATGAGTGGATAGTGGG 15636


RESULT 95
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             3.9%;  Score 38.6;  DB 3;  Length 4403765;
  Best Local Similarity   45.7%;  Pred. No. 4.9;
  Matches  209;  Conservative    0;  Mismatches  244;  Indels    4;  Gaps    2;

Qy        416 TGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCT 475
              ||||| || | | || |  | |        || || | ||||   | |       |  |||
Db    3925117 TGCTGGGGCGCCCGTTGATCGGTGACGGCGCCAATGCGACGACTCCCGGCGGGGCCGGCG 3925176

Qy        476 TGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG--CTGG 533
              |   |  || ||||||  | || |    ||   | |||| ||  |   | | |  | ||
Db    3925177 GGGCCGGCGGTCTGCTGTTCGGCAACGGCGGGGCCGGGGCAGCCGGGGCGCCCGGCCAGG 3925236
```

```
Qy        534 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGG 593
              | ||  |   ||| |    || |    |||     ||   ||  |  |||     |
Db    3925237 CCGGCGGGCCTGGCGGGCCCGCCGGATTGTGGGGCAACGGCGGGCCCGGCGGGGCCGGCG 3925296

Qy        594 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 653
              || ||     |||| || | | |||  |  | || | ||||  |||  | ||| |    |
Db    3925297 GCAGCGGTGGGGGCACCGGCGGTGCCGGCGGCGCCGGTGGGTGGCTGTTCGGGGTTGGCG 3925356

Qy        654 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCT--GCTGGCTTGCTTGCTTGCT 711
              ||  |      | | |  |||| | | |  | | | |   || |||  ||  | | | |
Db    3925357 GCGCCGGCGGTGTCAGTGGGGCCGGTGGCGGCACCGGCGGGGCGGGCGGGCCCGGTGGTT 3925416

Qy        712 TGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGG 771
              || |   |||       |    |  |  | | |  | || | |||  | || |    ||
Db    3925417 TGATCTGGGGCGGCGGCGGGGCCGGCGGTGTCGGTGGGGCCGGTGGCGGCACCGGCGGGG 3925476

Qy        772 GCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACCCGGGC 831
              | || | |  || | |     |  || |   ||    |  |  |   ||||  |||
Db    3925477 CCGGCGGCCGCGCCGAGCTGCTGTTCGGCGCCGGCGGTGCGGGTGGGGCGGGCACCGACG 3925536

Qy        832 CTCTCCCCCAGGCACAAAGCAGTAGGGCTGGCCAGGG 868
                  |||   ||     || | |  |  || || ||
Db    3925537 GCGGGCCCGGTGCTACCGGCGGGACCGGCGGACACGG 3925573


RESULT 96
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             3.9%;  Score 38.6;  DB 3;  Length 4411529;
  Best Local Similarity   45.7%;  Pred. No. 4.9;
  Matches  209;  Conservative    0;  Mismatches  244;  Indels    4;  Gaps    2;

Qy        416 TGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCT 475
              ||||| || | | || |  | |       || || | ||||   | |       |  |||
Db    3931329 TGCTGGGGCGCCCGTTGATCGGTGACGGCGCCAATGCGACGACTCCCGGCGGGGCCGGCG 3931388

Qy        476 TGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG--CTGG 533
              |   |  || |||||  | || |    ||   | |||| || |   | | |  | ||
Db    3931389 GGGCCGGCGGTCTGCTGTTCGGCAACGGCGGGGCCGGGGCAGCCGGGGCGCCCGGCCAGG 3931448

Qy        534 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGG 593
              | ||  |   ||| |    || |    |||     ||   ||  |  |||     |
Db    3931449 CCGGCGGGCCTGGCGGGCCCGCCGGATTGTGGGGCAACGGCGGGCCCGGCGGGGCCGGCG 3931508

Qy        594 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 653
              || ||     |||| || | | |||  |  | || | ||||  |||  | ||| |    |
Db    3931509 GCAGCGGTGGGGGCACCGGCGGTGCCGGCGGCGCCGGTGGGTGGCTGTTCGGGGTTGGCG 3931568

Qy        654 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCT--GCTGGCTTGCTTGCTTGCT 711
              ||  |      | | |  |||| | | |  | | | |   || |||  ||  | | | |
Db    3931569 GCGCCGGCGGTGTCGGTGGGGCCGGTGGCGGCACCGGCGGGGCGGGCGGGCCCGGTGGTT 3931628

Qy        712 TGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGG 771
              || |   |||       |    |  |  | | |  | || | |||  | || |    ||
Db    3931629 TGATCTGGGGCGGCGGCGGGGCCGGCGGTGTCGGTGGGGCCGGTGGCGGCACCGGCGGGG 3931688

Qy        772 GCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACCCGGGC 831
              | || | |  || | |     |  || |   ||    |  |  |   ||||  |||
Db    3931689 CCGGCGGCCGCGCCGAGCTGCTGTTCGGCGCCGGCGGTGCGGGTGGGGCGGGCACCGACG 3931748

Qy        832 CTCTCCCCCAGGCACAAAGCAGTAGGGCTGGCCAGGG 868
                  |||   ||     || | |  |  || || ||
Db    3931749 GCGGGCCCGGTGCTACCGGCGGGACCGGCGGACACGG 3931785


RESULT 97
US-07-922-723A-9/c
; Sequence 9, Application US/07922723A
; Patent No. 5369004
;  GENERAL INFORMATION:
;    APPLICANT:  Drs. Mihael H. Polymeropoulos
;    APPLICANT:  and Carl R. Merril
;    TITLE OF INVENTION:  FIVE HIGHLY INFORMATIVE
;    TITLE OF INVENTION:  REPEAT POLYMORPHIC DNA MARKERS
;    NUMBER OF SEQUENCES:  73
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lowe, Price, LeBlanc & Becker
;      STREET:  Suite 300, 99 Canal Center Plaza
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  DOS Text File
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/922,723A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  D.J. Mills
;      REGISTRATION NUMBER:  34506
;      REFERENCE/DOCKET NUMBER:  717081B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703 684 1111
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  243
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-07-922-723A-9

  Query Match             3.8%;  Score 38.2;  DB 1;  Length 243;
  Best Local Similarity   52.9%;  Pred. No. 0.18;
  Matches   82;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        253 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 312
              | || ||| ||  ||  ||| ||| ||| ||  ||  | | ||| ||| ||  ||  | |
Db        189 GTTTCCTTTCTTTCTTTCCTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT 130

Qy        313 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 372
              ||| ||| ||  ||  |   ||  ||  ||| ||  |    |   |   ||  |
Db        129 TCTTTCTTTCTTTCTTTCTCTCTCTCTCTCTCTCTCTCTTTCTTTCTTTCTTTCTTTC 70

Qy        373 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 407
                |   | |||  || | | ||| ||| || | |
Db         69 TTTCTTTCTCTCTTTGTCTTTCTTTCTTTCTTGTT 35


RESULT 98
US-07-799-828C-9/c
```

```
; Sequence 9, Application US/07799828C
; Patent No. 5378602
;  GENERAL INFORMATION:
;    APPLICANT:  Drs. Carl R. Merril and
;    APPLICANT:  Mihael H. Polymeropoulos
;    TITLE OF INVENTION:  TWENTY SEVEN HIGHLY INFORMATIVE
;    TITLE OF INVENTION:  MICROSATELLITE REPEAT
;    TITLE OF INVENTION:  POLYMORPHIC DNA MARKERS
;    NUMBER OF SEQUENCES:  63
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lowe, Price, LeBlanc & Becker
;      STREET:  Suite 300, 99 Canal Center Plaza
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  DOS Text File
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/799,828C
;      FILING DATE:  19911127
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  D.J. Mills
;      REGISTRATION NUMBER:  34,506
;      REFERENCE/DOCKET NUMBER:  717081A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703 684 1111
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  243
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-07-799-828C-9

  Query Match             3.8%;  Score 38.2;  DB 1;  Length 243;
  Best Local Similarity   52.9%;  Pred. No. 0.18;
  Matches   82;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        253 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 312
              | || ||| ||  ||  ||| ||| ||| ||  ||  | | ||| ||| ||  ||  | |
Db        189 GTTTCCTTTCTTTCTTTCCTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT 130

Qy        313 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 372
              ||| ||| ||  ||  |   ||  ||  ||| ||  |    |   |   ||  |
Db        129 TCTTTCTTTCTTTCTTTCTCTCTCTCTCTCTCTCTCTCTTTCTTTCTTTCTTTCTTTC 70

Qy        373 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 407
                |   |  |||  || | | ||| ||| || | |
Db         69 TTTCTTTCTCTCTTTGTCTTTCTTTCTTTCTTGTT 35


RESULT 99
US-08-074-275-9/c
; Sequence 9, Application US/08074275
; Patent No. 5468610
;  GENERAL INFORMATION:
;    APPLICANT:  Drs. Carl R. Merril and
;    APPLICANT:  Mihael H. Polymeropoulos
;    TITLE OF INVENTION:  THREE HIGHLY INFORMATIVE REPEAT
;    TITLE OF INVENTION:  POLYMORPHIC DNA MARKERS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lowe, Price, LeBlanc & Becker
;      STREET:  Suite 300, 99 Canal Center Plaza
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  DOS Text File
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/074,275
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/707,501
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  J.G. Mullins
;      REGISTRATION NUMBER:  33073
;      REFERENCE/DOCKET NUMBER:  717081
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703 684 1111
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  243
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-074-275-9

  Query Match             3.8%;  Score 38.2;  DB 1;  Length 243;
  Best Local Similarity   52.9%;  Pred. No. 0.18;
  Matches   82;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        253 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 312
              | || ||| ||  ||  ||| ||| ||| ||  ||  | | ||| ||| ||  ||  | |
Db        189 GTTTCCTTTCTTTCTTTCCTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT 130

Qy        313 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 372
              ||| ||| ||  ||  |   ||  ||  ||| ||  |    |   |   ||  |
Db        129 TCTTTCTTTCTTTCTTTCTCTCTCTCTCTCTCTCTCTCTTTCTTTCTTTCTTTCTTTC 70

Qy        373 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 407
                |   |  |||  || | | ||| ||| || | |
Db         69 TTTCTTTCTCTCTTTGTCTTTCTTTCTTTCTTGTT 35


RESULT 100
US-08-480-366-9/c
; Sequence 9, Application US/08480366
; Patent No. 5721100
;  GENERAL INFORMATION:
;    APPLICANT:  Drs. Carl R. Merril and
;    APPLICANT:  Mihael H. Polymeropoulos
;    TITLE OF INVENTION:  THREE HIGHLY INFORMATIVE REPEAT
;    TITLE OF INVENTION:  POLYMORPHIC DNA MARKERS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lowe, Price, LeBlanc & Becker
;      STREET:  Suite 300, 99 Canal Center Plaza
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  DOS Text File
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,366
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  J.G. Mullins
```

```
;     REGISTRATION NUMBER:  33073
;     REFERENCE/DOCKET NUMBER:  717081
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  703 684 1111
;  INFORMATION FOR SEQ ID NO:  9:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  243
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  DNA (genomic)
US-08-480-366-9

  Query Match             3.8%;  Score 38.2;  DB 1;  Length 243;
  Best Local Similarity   52.9%;  Pred. No. 0.18;
  Matches   82;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        253 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 312
              | || ||| ||  ||  ||| ||| ||| ||  ||  | | ||| ||| ||  ||  | |
Db        189 GTTTCCTTTCTTTCTTTCCTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT 130

Qy        313 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 372
               ||| ||| ||  ||  |  || ||  ||| ||  |    |   |   ||  |
Db        129 TCTTTCTTTCTTTCTTTCTCTCTCTCTCTCTCTCTCTCTCTTTCTTTCTTTCTTTCTTTC 70

Qy        373 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 407
                |   |  |||  || | | ||| ||| || | |
Db         69 TTTCTTTCTCTCTTTGTCTTTCTTTCTTTCTTGTT 35


Search completed: August 22, 2003, 00:41:55
Job time : 119.962 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 21, 2003, 21:15:12 ; Search time 246.377 Seconds
(without alignments)
9124.900 Million cell updates/sec

Title: US-09-831-000-1_COPY_1_1000
Perfect score: 1000
Sequence: 1 gatcgggaaaacgcgagggg.....gtgcacggccaagttacaag 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1517243 seqs, 1124081882 residues

Total number of hits satisfying chosen parameters: 3034486

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	192.8	19.3	576	9	US-09-864-761-26582	Sequence 26582, A
	2	180.6	18.1	555	9	US-09-864-761-9605	Sequence 9605, Ap
	3	117.8	11.8	1360	12	US-10-017-161-1345	Sequence 1345, Ap
	4	94	9.4	1672	12	US-10-017-161-2117	Sequence 2117, Ap
c	5	92.2	9.2	798	13	US-10-027-632-164203	Sequence 164203,
c	6	92.2	9.2	798	13	US-10-027-632-164204	Sequence 164204,
	7	83.6	8.4	14750	12	US-10-017-161-1915	Sequence 1915, Ap
c	8	82.6	8.3	572	13	US-10-027-632-272985	Sequence 272985,
	9	80.8	8.1	42999	10	US-09-799-462A-17	Sequence 17, Appl
	10	80.8	8.1	42999	11	US-09-836-911A-17	Sequence 17, Appl
	11	80.8	8.1	42999	13	US-10-125-767-17	Sequence 17, Appl
	12	80.8	8.1	42999	14	US-10-151-081-17	Sequence 17, Appl
	13	80.8	8.1	42999	14	US-10-287-313-17	Sequence 17, Appl
	14	80.8	8.1	42999	14	US-10-219-694-17	Sequence 17, Appl
	15	80	8.0	1297	12	US-10-017-161-2193	Sequence 2193, Ap
	16	79.4	7.9	3213	13	US-10-027-632-113494	Sequence 113494,
c	17	78.4	7.8	388	9	US-09-728-446-295	Sequence 295, App
c	18	78.4	7.8	428	9	US-09-728-446-369	Sequence 369, App
c	19	78.4	7.8	436	9	US-09-728-446-354	Sequence 354, App
c	20	77.4	7.7	409	9	US-09-728-446-362	Sequence 362, App
	21	77.4	7.7	1830121	14	US-10-329-960-1	Sequence 1, Appli
	22	77.2	7.7	368	11	US-09-764-891-8904	Sequence 8904, Ap
	23	77.2	7.7	368	11	US-09-764-891-10047	Sequence 10047, A
c	24	77.2	7.7	392	9	US-09-728-446-736	Sequence 736, App
c	25	75.8	7.6	410	9	US-09-728-446-325	Sequence 325, App
c	26	74.4	7.4	427	9	US-09-728-446-250	Sequence 250, App
c	27	73.8	7.4	1006	13	US-10-027-632-273001	Sequence 273001,
c	28	72.8	7.3	379	9	US-09-728-446-246	Sequence 246, App
c	29	72.6	7.3	368	13	US-10-040-739-1102	Sequence 1102, Ap
c	30	69.6	7.0	681	13	US-10-027-632-141458	Sequence 141458,
c	31	68.4	6.8	681	13	US-10-027-632-141457	Sequence 141457,
c	32	68.4	6.8	681	13	US-10-027-632-141460	Sequence 141460,
c	33	68	6.8	681	13	US-10-027-632-141459	Sequence 141459,
	34	67	6.7	397	9	US-09-728-446-465	Sequence 465, App
	35	65.8	6.6	223	10	US-09-764-877-3185	Sequence 3185, Ap
c	36	65.2	6.5	393	9	US-09-728-446-978	Sequence 978, App
	37	65.2	6.5	13830	12	US-10-017-161-1581	Sequence 1581, Ap
	38	65.2	6.5	31718	11	US-09-764-872-812	Sequence 812, App
	39	65.2	6.5	31718	11	US-09-764-872-813	Sequence 813, App
	40	65.2	6.5	31718	11	US-09-764-891-9103	Sequence 9103, Ap
	41	65.2	6.5	31718	11	US-09-764-891-9104	Sequence 9104, Ap
c	42	65	6.5	647	13	US-10-027-632-281951	Sequence 281951,
	43	64.6	6.5	495	13	US-10-027-632-273402	Sequence 273402,
	44	64.4	6.4	90650	14	US-10-175-523-80	Sequence 80, Appl
	45	64.2	6.4	223	10	US-09-764-877-3186	Sequence 3186, Ap
c	46	63.4	6.3	3595	10	US-09-089-818B-6	Sequence 6, Appli
	47	63.2	6.3	863	13	US-10-027-632-161036	Sequence 161036,
	48	63.2	6.3	863	13	US-10-027-632-161037	Sequence 161037,
	49	63.2	6.3	863	13	US-10-027-632-161038	Sequence 161038,
	50	62.8	6.3	6109	12	US-10-311-455-300	Sequence 300, App
	51	62.8	6.3	110079	14	US-10-175-523-96	Sequence 96, Appl
c	52	62.2	6.2	2940917	13	US-10-027-632-174763	Sequence 174763,
	53	62	6.2	726	13	US-10-027-632-15974	Sequence 15974, A
	54	62	6.2	13240	12	US-10-017-161-1673	Sequence 1673, Ap
c	55	61.8	6.2	1771	9	US-09-874-923-7	Sequence 7, Appli
c	56	61.8	6.2	1771	10	US-09-991-496-7	Sequence 7, Appli
	57	60.6	6.1	588	13	US-10-027-632-143381	Sequence 143381,
c	58	60.6	6.1	15720	9	US-09-922-217-1058	Sequence 1058, Ap
c	59	60.6	6.1	15720	10	US-09-833-263-1058	Sequence 1058, Ap
c	60	60.6	6.1	15720	13	US-10-025-380-1058	Sequence 1058, Ap
c	61	60.4	6.0	6171	12	US-10-311-455-762	Sequence 762, App
	62	60.2	6.0	588	13	US-10-027-632-143382	Sequence 143382,
	63	60.2	6.0	588	13	US-10-027-632-143383	Sequence 143383,
	64	59	5.9	553	13	US-10-027-632-183437	Sequence 183437,
c	65	58.8	5.9	159095	14	US-10-017-128-3	Sequence 3, Appli
	66	58.6	5.9	300000	14	US-10-262-552-33	Sequence 33, Appl
	67	58.2	5.8	489	13	US-10-027-632-170164	Sequence 170164,
	68	58.2	5.8	489	13	US-10-027-632-170165	Sequence 170165,
	69	58.2	5.8	978	13	US-10-027-632-170163	Sequence 170163,
c	70	57.6	5.8	439	13	US-10-027-632-68339	Sequence 68339, A
c	71	57.6	5.8	439	13	US-10-027-632-68340	Sequence 68340, A
c	72	57.6	5.8	439	13	US-10-027-632-295071	Sequence 295071,
c	73	57.6	5.8	439	13	US-10-027-632-295072	Sequence 295072,
c	74	57.6	5.8	549	13	US-10-027-632-37032	Sequence 37032, A
c	75	57.6	5.8	549	13	US-10-027-632-37033	Sequence 37033, A
c	76	56.6	5.7	412	11	US-09-803-719-1534	Sequence 1534, Ap
	77	56.2	5.6	170834	9	US-09-835-232-7	Sequence 7, Appli
c	78	56	5.6	573	9	US-09-864-761-9961	Sequence 9961, Ap
c	79	56	5.6	573	9	US-09-864-761-26463	Sequence 26463, A
c	80	55.2	5.5	502	13	US-10-027-632-270367	Sequence 270367,
c	81	55.2	5.5	502	13	US-10-027-632-270368	Sequence 270368,
	82	55.2	5.5	631	13	US-10-027-632-174148	Sequence 174148,
	83	55.2	5.5	631	13	US-10-027-632-174149	Sequence 174149,
c	84	54.8	5.5	345	9	US-09-728-446-190	Sequence 190, App
c	85	54.8	5.5	12007	12	US-10-311-455-689	Sequence 689, App
	86	54.8	5.5	85548	14	US-10-175-523-75	Sequence 75, Appl
	87	54.8	5.5	1223197	13	US-10-027-632-179264	Sequence 179264,
c	88	54.6	5.5	354	9	US-09-728-446-435	Sequence 435, App
c	89	54.6	5.5	707	13	US-10-027-632-141461	Sequence 141461,

c	90	54	5.4	589	13	US-10-027-632-86169	Sequence 86169, A
c	91	54	5.4	225883	14	US-10-175-523-57	Sequence 57, Appl
c	92	53.8	5.4	2201	12	US-09-791-254-4	Sequence 4, Appli
c	93	53.8	5.4	4065	12	US-09-791-254-1	Sequence 1, Appli
c	94	53.8	5.4	1223197	13	US-10-027-632-179264	Sequence 179264,
	95	53.6	5.4	26555	10	US-09-860-670-161	Sequence 161, App
c	96	53.6	5.4	30000	14	US-10-007-010-10	Sequence 10, Appl
	97	53.6	5.4	136328	12	US-10-101-510-127	Sequence 127, App
	98	53.4	5.3	1297	12	US-10-017-161-2377	Sequence 2377, Ap
	99	53.2	5.3	766	13	US-10-027-632-30729	Sequence 30729, A
	100	53.2	5.3	766	13	US-10-027-632-30730	Sequence 30730, A
c	101	52.6	5.3	397658	10	US-09-813-320-3	Sequence 3, Appli
c	102	52.2	5.2	582	13	US-10-027-632-269946	Sequence 269946,
	103	52	5.2	573	9	US-09-864-761-9961	Sequence 9961, Ap
	104	52	5.2	573	9	US-09-864-761-26463	Sequence 26463, A
c	105	51.6	5.2	636	13	US-10-027-632-308921	Sequence 308921,
c	106	51.6	5.2	636	13	US-10-027-632-308922	Sequence 308922,
c	107	51.4	5.1	155074	13	US-10-026-188-6	Sequence 6, Appli
	108	51.2	5.1	486	13	US-10-027-632-135836	Sequence 135836,
	109	51.2	5.1	2487	13	US-10-027-632-103365	Sequence 103365,
c	110	51	5.1	369	10	US-09-867-701-5268	Sequence 5268, Ap
	111	51	5.1	605	13	US-10-027-632-271134	Sequence 271134,
c	112	51	5.1	672	13	US-10-027-632-106237	Sequence 106237,
c	113	50.8	5.1	263744	12	US-10-229-834A-6	Sequence 6, Appli
	114	50.8	5.1	659158	10	US-09-771-208-20	Sequence 20, Appl
c	115	50.6	5.1	672	13	US-10-027-632-137449	Sequence 137449,
c	116	50.6	5.1	672	13	US-10-027-632-137450	Sequence 137450,
c	117	50.6	5.1	1360	12	US-10-017-161-1345	Sequence 1345, Ap
	118	50.4	5.0	702	13	US-10-027-632-110828	Sequence 110828,
	119	50.4	5.0	702	13	US-10-027-632-110829	Sequence 110829,
	120	50.4	5.0	702	13	US-10-027-632-110830	Sequence 110830,
c	121	50.4	5.0	1682	9	US-09-755-325-7	Sequence 7, Appli
c	122	50.4	5.0	8367	12	US-10-017-161-807	Sequence 807, App
c	123	50.2	5.0	1446	14	US-10-156-761-4465	Sequence 4465, Ap
c	124	50.2	5.0	2000	10	US-09-938-842A-5032	Sequence 5032, Ap
	125	50.2	5.0	9025608	14	US-10-156-761-1	Sequence 1, Appli
c	126	50	5.0	577	13	US-10-027-632-231482	Sequence 231482,
	127	50	5.0	22400	14	US-10-143-266-1	Sequence 1, Appli
	128	49.8	5.0	745	13	US-10-027-632-144367	Sequence 144367,
	129	49.6	5.0	480	13	US-10-027-632-72218	Sequence 72218, A
	130	49.6	5.0	480	13	US-10-027-632-72219	Sequence 72219, A
	131	49.6	5.0	480	13	US-10-027-632-72220	Sequence 72220, A
	132	49.6	5.0	480	13	US-10-027-632-72801	Sequence 72801, A
	133	49.6	5.0	480	13	US-10-027-632-72802	Sequence 72802, A
	134	49.6	5.0	480	13	US-10-027-632-108633	Sequence 108633,
	135	49.6	5.0	480	13	US-10-027-632-108634	Sequence 108634,
	136	49.6	5.0	480	13	US-10-027-632-108635	Sequence 108635,
	137	49.6	5.0	7597	12	US-10-311-455-985	Sequence 985, App
	138	49.6	5.0	90541	9	US-09-759-359A-3	Sequence 3, Appli
c	139	49.4	4.9	519	9	US-09-864-761-12571	Sequence 12571, A
	140	49.4	4.9	931	13	US-10-027-632-120060	Sequence 120060,
	141	49.4	4.9	931	13	US-10-027-632-120061	Sequence 120061,
	142	49.4	4.9	931	13	US-10-027-632-120062	Sequence 120062,
	143	49.4	4.9	3593	13	US-10-002-775-10	Sequence 10, Appl
	144	49.4	4.9	3593	14	US-10-115-615-22	Sequence 22, Appl
	145	49.4	4.9	12237	12	US-10-311-455-2332	Sequence 2332, Ap
	146	49.2	4.9	572	13	US-10-027-632-321875	Sequence 321875,
	147	49.2	4.9	702	13	US-10-027-632-126555	Sequence 126555,
	148	49.2	4.9	249487	13	US-10-026-188-3	Sequence 3, Appli
	149	49	4.9	157875	11	US-09-935-464-1	Sequence 1, Appli
	150	49	4.9	157875	14	US-10-125-835-1	Sequence 1, Appli
c	151	48.8	4.9	210	10	US-09-933-797-404	Sequence 404, App
	152	48.8	4.9	239	13	US-10-027-632-283353	Sequence 283353,
	153	48.8	4.9	499	13	US-10-027-632-195772	Sequence 195772,
c	154	48.8	4.9	2892	14	US-10-175-523-103	Sequence 103, App
c	155	48.6	4.9	1973	13	US-10-027-632-97352	Sequence 97352, A
	156	48.6	4.9	3503	9	US-09-823-038A-58	Sequence 58, Appl
c	157	48.4	4.8	585	13	US-10-027-632-319580	Sequence 319580,
c	158	48.4	4.8	585	13	US-10-027-632-319581	Sequence 319581,
c	159	48.4	4.8	585	13	US-10-027-632-319582	Sequence 319582,
c	160	48.4	4.8	3038	13	US-10-027-632-115184	Sequence 115184,
c	161	48.2	4.8	598	13	US-10-027-632-226940	Sequence 226940,
c	162	48	4.8	15140	12	US-10-017-161-1101	Sequence 1101, Ap
c	163	48	4.8	15182	12	US-10-017-161-1541	Sequence 1541, Ap
c	164	47.8	4.8	516	13	US-10-027-632-41615	Sequence 41615, A
c	165	47.8	4.8	561	13	US-10-027-632-77276	Sequence 77276, A
	166	47.8	4.8	953	13	US-10-027-632-165095	Sequence 165095,
	167	47.8	4.8	3133	12	US-10-017-161-1483	Sequence 1483, Ap
c	168	47.6	4.8	607	13	US-10-027-632-42357	Sequence 42357, A
c	169	47.6	4.8	607	13	US-10-027-632-42358	Sequence 42358, A
c	170	47.6	4.8	607	13	US-10-027-632-63776	Sequence 63776, A
c	171	47.6	4.8	607	13	US-10-027-632-63777	Sequence 63777, A
c	172	47.6	4.8	16554	12	US-10-017-161-1961	Sequence 1961, Ap
	173	47.4	4.7	6419	12	US-10-311-455-239	Sequence 239, App
c	174	47.2	4.7	576	13	US-10-027-632-272310	Sequence 272310,
	175	47	4.7	628	13	US-10-027-632-11159	Sequence 11159, A
	176	47	4.7	628	13	US-10-027-632-11160	Sequence 11160, A
	177	47	4.7	628	13	US-10-027-632-143978	Sequence 143978,
c	178	47	4.7	2723	13	US-10-027-632-112337	Sequence 112337,
	179	47	4.7	14776	10	US-09-880-107-2088	Sequence 2088, Ap
c	180	47	4.7	18443	11	US-09-728-552-6	Sequence 6, Appli
	181	47	4.7	130427	14	US-10-175-523-87	Sequence 87, Appl
	182	47	4.7	1691139	14	US-10-067-514-1	Sequence 1, Appli
	183	46.8	4.7	666	13	US-10-027-632-28781	Sequence 28781, A
c	184	46.8	4.7	1166	13	US-10-027-632-251519	Sequence 251519,
c	185	46.8	4.7	1166	13	US-10-027-632-251520	Sequence 251520,
	186	46.8	4.7	38186	11	US-09-373-658-38	Sequence 38, Appl
	187	46.6	4.7	597	13	US-10-027-632-40862	Sequence 40862, A
	188	46.6	4.7	633	13	US-10-027-632-61406	Sequence 61406, A
	189	46.6	4.7	633	13	US-10-027-632-297806	Sequence 297806,
c	190	46.4	4.6	702	13	US-10-027-632-201205	Sequence 201205,
c	191	46.4	4.6	2000	10	US-09-887-576-54	Sequence 54, Appl
c	192	46.4	4.6	11729	12	US-10-311-455-867	Sequence 867, App
c	193	46.4	4.6	2140405	13	US-10-027-632-76212	Sequence 76212, A
c	194	46.2	4.6	564	13	US-10-027-632-274098	Sequence 274098,
c	195	46.2	4.6	693	10	US-09-764-877-2421	Sequence 2421, Ap
	196	46.2	4.6	1218	13	US-10-027-632-256923	Sequence 256923,
	197	46.2	4.6	2436	13	US-10-027-632-256922	Sequence 256922,
	198	46.2	4.6	110096	10	US-09-880-107-1542	Sequence 1542, Ap
c	199	46.2	4.6	152331	13	US-10-095-407-16	Sequence 16, Appl
	200	46.2	4.6	176373	13	US-10-095-407-17	Sequence 17, Appl
	201	46	4.6	35425	12	US-10-017-161-2429	Sequence 2429, Ap
c	202	45.8	4.6	563	13	US-10-027-632-210174	Sequence 210174,
c	203	45.8	4.6	563	13	US-10-027-632-210175	Sequence 210175,
c	204	45.8	4.6	32082	11	US-09-764-891-9679	Sequence 9679, Ap
c	205	45.8	4.6	145831	10	US-09-969-708-79	Sequence 79, Appl
c	206	45.8	4.6	145831	10	US-09-954-456-2116	Sequence 2116, Ap
c	207	45.6	4.6	732	13	US-10-027-632-21127	Sequence 21127, A
c	208	45.6	4.6	90650	14	US-10-175-523-80	Sequence 80, Appl
c	209	45.2	4.5	496	13	US-10-013-329-9	Sequence 9, Appli
c	210	45.2	4.5	771	13	US-10-027-632-133833	Sequence 133833,
c	211	45.2	4.5	771	13	US-10-027-632-133834	Sequence 133834,
c	212	45.2	4.5	819	13	US-10-027-632-130210	Sequence 130210,
c	213	45.2	4.5	2264	13	US-10-027-632-111056	Sequence 111056,
	214	45	4.5	286	10	US-09-867-701-8693	Sequence 8693, Ap
c	215	45	4.5	564	13	US-10-027-632-274099	Sequence 274099,
	216	45	4.5	792	13	US-10-027-632-168583	Sequence 168583,
	217	45	4.5	792	13	US-10-027-632-168584	Sequence 168584,
	218	45	4.5	792	13	US-10-027-632-168585	Sequence 168585,
	219	45	4.5	792	13	US-10-027-632-168586	Sequence 168586,
	220	45	4.5	832	13	US-10-027-632-170046	Sequence 170046,
	221	45	4.5	832	13	US-10-027-632-170047	Sequence 170047,
	222	45	4.5	832	13	US-10-027-632-170048	Sequence 170048,
c	223	45	4.5	861	13	US-10-027-632-154904	Sequence 154904,
c	224	45	4.5	1647	13	US-10-027-632-154905	Sequence 154905,
	225	45	4.5	1747	12	US-10-017-161-2371	Sequence 2371, Ap
	226	45	4.5	1747	12	US-10-017-161-2379	Sequence 2379, Ap
c	227	45	4.5	2848	13	US-10-027-632-112062	Sequence 112062,
c	228	45	4.5	123192	14	US-10-175-523-71	Sequence 71, Appl
	229	45	4.5	173808	13	US-10-003-806-10	Sequence 10, Appl
	230	45	4.5	1601042	13	US-10-027-632-59064	Sequence 59064, A
	231	44.8	4.5	468	11	US-09-918-995-11647	Sequence 11647, A
	232	44.8	4.5	920	13	US-10-027-632-323863	Sequence 323863,
c	233	44.6	4.5	1220	13	US-10-027-632-123721	Sequence 123721,
c	234	44.6	4.5	1220	13	US-10-027-632-123722	Sequence 123722,
c	235	44.6	4.5	1424	9	US-09-729-674-71	Sequence 71, Appl

c 236	44.4	4.4	362	13	US-10-027-632-285908	Sequence 285908,
c 237	44.4	4.4	1854	10	US-09-912-599-2	Sequence 2, Appli
c 238	44.4	4.4	12138	12	US-10-311-455-1602	Sequence 1602, Ap
c 239	44.4	4.4	186510	12	US-10-043-715-1	Sequence 1, Appli
c 240	44.2	4.4	577	13	US-10-027-632-267218	Sequence 267218,
c 241	44.2	4.4	19994	12	US-10-017-161-729	Sequence 729, App
c 242	44	4.4	506	13	US-10-027-632-196339	Sequence 196339,
243	44	4.4	652	13	US-10-027-632-247562	Sequence 247562,
244	44	4.4	652	13	US-10-027-632-247563	Sequence 247563,
245	44	4.4	652	13	US-10-027-632-247564	Sequence 247564,
c 246	44	4.4	702	13	US-10-027-632-110828	Sequence 110828,
c 247	44	4.4	702	13	US-10-027-632-110829	Sequence 110829,
c 248	44	4.4	702	13	US-10-027-632-110830	Sequence 110830,
249	44	4.4	777	13	US-10-027-632-160333	Sequence 160333,
c 250	44	4.4	33719	15	US-10-080-170-651	Sequence 651, App
c 251	43.8	4.4	20966	9	US-09-776-976-7	Sequence 7, Appli
c 252	43.8	4.4	20966	9	US-09-758-055-7	Sequence 7, Appli
c 253	43.8	4.4	20966	9	US-09-909-547-7	Sequence 7, Appli
c 254	43.8	4.4	20966	14	US-10-231-814-7	Sequence 7, Appli
c 255	43.6	4.4	299	10	US-09-867-701-9477	Sequence 9477, Ap
c 256	43.4	4.3	798	13	US-10-027-632-30451	Sequence 30451, A
257	43.4	4.3	14221	11	US-09-764-891-9257	Sequence 9257, Ap
258	43.4	4.3	14221	14	US-10-091-572-833	Sequence 833, App
259	43.4	4.3	52216	9	US-09-747-810-1	Sequence 1, Appli
c 260	43.2	4.3	575	13	US-10-027-632-50030	Sequence 50030, A
c 261	43.2	4.3	575	13	US-10-027-632-81841	Sequence 81841, A
c 262	43.2	4.3	575	13	US-10-027-632-302015	Sequence 302015,
263	43.2	4.3	759	13	US-10-027-632-110083	Sequence 110083,
264	43.2	4.3	174424	10	US-09-967-768A-314	Sequence 314, App
265	43.2	4.3	174424	12	US-09-960-706-969	Sequence 969, App
c 266	43	4.3	307	10	US-09-867-701-9416	Sequence 9416, Ap
c 267	43	4.3	725	13	US-10-027-632-251098	Sequence 251098,
c 268	43	4.3	725	13	US-10-027-632-251099	Sequence 251099,
c 269	43	4.3	733	13	US-10-027-632-29177	Sequence 29177, A
c 270	43	4.3	733	13	US-10-027-632-29178	Sequence 29178, A
271	43	4.3	2933	13	US-10-027-632-112347	Sequence 112347,
c 272	43	4.3	39788	12	US-10-017-161-719	Sequence 719, App
c 273	42.8	4.3	362	13	US-10-027-632-285907	Sequence 285907,
c 274	42.8	4.3	488	13	US-10-027-632-307123	Sequence 307123,
c 275	42.8	4.3	488	13	US-10-027-632-307124	Sequence 307124,
c 276	42.8	4.3	602	13	US-10-027-632-278226	Sequence 278226,
277	42.8	4.3	4223	9	US-09-737-178-7	Sequence 7, Appli
c 278	42.8	4.3	4223	9	US-09-737-178-45	Sequence 45, Appl
279	42.8	4.3	4223	10	US-09-286-488-7	Sequence 7, Appli
c 280	42.8	4.3	4223	10	US-09-286-488-45	Sequence 45, Appl
281	42.8	4.3	4223	11	US-09-853-079-7	Sequence 7, Appli
c 282	42.8	4.3	4223	11	US-09-853-079-45	Sequence 45, Appl
c 283	42.8	4.3	45000	14	US-10-000-213-12	Sequence 12, Appl
c 284	42.8	4.3	122186	11	US-09-563-728A-36	Sequence 36, Appl
c 285	42.8	4.3	249487	13	US-10-026-188-3	Sequence 3, Appli
c 286	42.6	4.3	535	13	US-10-027-632-275062	Sequence 275062,
287	42.6	4.3	577	13	US-10-027-632-267218	Sequence 267218,
c 288	42.6	4.3	577	13	US-10-027-632-267219	Sequence 267219,
c 289	42.6	4.3	684	13	US-10-027-632-21356	Sequence 21356, A
290	42.6	4.3	1117	12	US-10-017-161-1403	Sequence 1403, Ap
c 291	42.6	4.3	2450	14	US-10-028-392-1	Sequence 1, Appli
c 292	42.6	4.3	2756	14	US-10-028-392-34	Sequence 34, Appl
c 293	42.6	4.3	19846	9	US-09-764-869-1682	Sequence 1682, Ap
c 294	42.6	4.3	19846	9	US-09-764-869-1683	Sequence 1683, Ap
c 295	42.6	4.3	19846	14	US-10-091-504-1682	Sequence 1682, Ap
c 296	42.6	4.3	19846	14	US-10-091-504-1683	Sequence 1683, Ap
c 297	42.4	4.2	683	13	US-10-027-632-137448	Sequence 137448,
298	42.4	4.2	5452	12	US-10-017-161-1481	Sequence 1481, Ap
c 299	42.4	4.2	6115	12	US-10-311-455-1773	Sequence 1773, Ap
c 300	42.4	4.2	13830	12	US-10-017-161-1581	Sequence 1581, Ap

ALIGNMENTS

RESULT 1
US-09-864-761-26582
; Sequence 26582, Application US/09864761

; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 26582
; LENGTH: 576
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000111.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
; OTHER INFORMATION: SWISSPROT HIT: P44836, EVALUE 1.00e-12
US-09-864-761-26582

Query Match 19.3%; Score 192.8; DB 9; Length 576;
Best Local Similarity 59.2%; Pred. No. 2.6e-44;
Matches 329; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

```
Qy        180 TGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGC 239
              || ||| |   | |  |||||| |  |  |   |   |  || | |  || | |   ||
Db          8 TGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTTGGCTGGTTGCTTGGT 67

Qy        240 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 299
              || | ||| |  |||||| ||| | | | |  ||||||||||||| ||| || ||||||||
Db         68 TGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 127

Qy        300 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 359
```

```
Db        128 TGCTTGCTTGCTTGGCTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGCTTGCTTGCT 187

Qy        360 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419

Db        188 TGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCT 247

Qy        420 GAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGAT 479

Db        248 TCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGCT 307

Qy        480 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 539

Db        308 TGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGTTTGGTTGCTTGGTTGGT 367

Qy        540 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 599

Db        368 TGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTTGTTGCTTGCTTGGT 427

Qy        600 AGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 659

Db        428 TCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTCGTTGAT 487

Qy        660 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTG 719

Db        488 TGCTTGGTTGGTTGGTTGCTTGGTTGGTTGGCTGGTTCATTGGTTGGTTGGTTGCTTGGT 547

Qy        720 GGGCCGCTTGCCTGCT 735

Db        548 TGTTTGGTTGGTTGGT 563


RESULT 2
US-09-864-761-9605
; Sequence 9605, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 9605
;  LENGTH: 555
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  OTHER INFORMATION: MAP TO AP000111.1
;  OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
;  OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
US-09-864-761-9605

  Query Match           18.1%;  Score 180.6;  DB 9;  Length 555;
  Best Local Similarity  60.1%;  Pred. No. 6.5e-41;
  Matches  300;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy        217 GGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 276

Db         56 GGTTGCTTGGTTGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTTGGCT 115

Qy        277 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 336

Db        116 GGTTGCTTGGTTGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTGCTT 175

Qy        337 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 396

Db        176 GCTTGCTTGCTTGCTTGCTTGCTTGGCTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTT 235

Qy        397 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACG 456

Db        236 GCTTGCTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCATTGGTTGCTT 295

Qy        457 GTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 516

Db        296 GGTTGCTTGCTTCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTTGGTT 355

Qy        517 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCT 576

Db        356 GGTTGGTTGCTTGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGTTTGGTT 415

Qy        577 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCT 636

Db        416 GCTTGGTTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTTGTT 475

Qy        637 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTG 696

Db        476 GCTTGCTTGGTTCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 535

Qy        697 GCTTGCTTGCTTGCTTGCT 715

Db        536 GGTTCGTTGATTGCTTGGT 554


RESULT 3
US-10-017-161-1345
; Sequence 1345, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
```

```
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1345
;   LENGTH: 1360
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1360)
;   FEATURE:  .
;   NAME/KEY: CDS
;   LOCATION: (201)..(1160)
US-10-017-161-1345

  Query Match            11.8%;  Score 117.8;  DB 12;  Length 1360;
  Best Local Similarity   50.2%;  Pred. No. 2.4e-23;
  Matches  318;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy         69 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 128
              | ||||||||||||  ||  |||||| |||       | | ||  ||| |  |||  ||
Db        328 GTCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTGTCTGCCCGTCTGCCTGCCC 387

Qy        129 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 188
                |||| ||       | | ||  | |||  ||   |   | |  |  ||||
Db        388 ATCTGCCTGTCTGTCTGCCTGTCCGTCTGTCTGTCCATCTGTCCATCTGCCTATCCATCT 447

Qy        189 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 248
              ||  ||||  |||       | | |   | |||   || ||     |  | |      ||
Db        448 GCCTGCCTGTCTGCCTGTCCGTCTGCCTGTCTGTCTGCCTGTCCATCTGTCCATCTGCCT 507

Qy        249 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 308
                |   ||| |||  |   | | || ||| |||  ||  | |||  ||||  ||  |||
Db        508 ATCCATCTGCCTGCCTGTCTGTCGGCCTGCCTGCCTGCCTGTCTGTCTGCTGCCTGTCTG 567

Qy        309 ---GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
                | |||| ||  |||  |   | |||| ||| ||  ||   |||||   |  |||  |
Db        568 TCCGTCTGCCTGTCTGCCTGTCCGTCTGCCTGCCTGTCCGTCTGCCTGTCCGTCTGCCTG 627

Qy        366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 425
              | |||||  ||  |||  ||  | |||| ||| ||  ||   |||||   |       |
Db        628 CCTGCCTGTCTGTCTGCCTGCCTGTCTGCCTGCCTGTCCGTCTGCCTGTCCGTCTGCCTG 687

Qy        426 ACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGG 485
                | ||  | |||| ||  |||       | | ||  ||||| |||   || ||   ||  |
Db        688 CCTGTCTGCCTGCCTGTCTGCCTGTCTGCCCGTCTGCCTGTCTGTCTGCCTGTCCGTCTG 747

Qy        486 GCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGG 545
                |||  | | |  ||   ||||   |     |  | ||   | |||   ||  |
Db        748 CCTGTCTGTCCGTCTGTCCATCTGCCTATCCATCTGCCTGCCTATCTGTCTGTCCGTCTG 807

Qy        546 GCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGG 605
                ||||  | |||   |   | || |  | |||   |   | |||     ||||   |
Db        808 CCTGCCTGTCTGTCTGCCTGTCTGCCTGTCTGTCTGCCTGTCTGTCCATCTGCCTATCCA 867

Qy        606 GCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGG 665
                || |  | ||||  ||  | |||   | ||||  ||  | ||||  | |||   ||  |
Db        868 TCTACCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTGTCTGTCTGCCTGCCTGTCTGTCTG 927

Qy        666 GCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGC 698
                |||   || |||| ||     || |  |  ||
Db        928 TCTGTCTGGTTGCTTGTGCATGTGTCCCCCAGC 960


RESULT 4
US-10-017-161-2117
; Sequence 2117, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2117
;   LENGTH: 1672
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1672)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1472)
US-10-017-161-2117

  Query Match             9.4%;  Score 94;  DB 12;  Length 1672;
  Best Local Similarity   47.9%;  Pred. No. 1.1e-16;
  Matches  301;  Conservative    0;  Mismatches  325;  Indels    2;  Gaps    1;

Qy         30 CAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGG 89
              ||||  |   | |  |||     |||| |    | ||| ||  |||  |||       |   |
Db        765 CAGGTGTGCGTGCTTGCAGCTGTGTGTGCTTGCAGGTGCGTGTGCTTGCAGCTGTGTGTG 824

Qy         90 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 149
              | |||  |     | |   | |||   |  |||| || |   ||  ||||    |    |
Db        825 CTTGCAGGTGTGTGTGCTTGCAGGTGCGTGTGCTTGCAGCTGTGTGTGCTTGCAGCTGTG 884

Qy        150 TAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 209
              |  || |||         | |   |  ||   |  |||| || |   ||  ||||    ||
Db        885 TGTGCTTGCAGCTGTGTGTGCTTGCTGGTGCGTGTGCTTGCAGCTGTGTGTGCTTGCAGG 944

Qy        210 ACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 269
                ||  |  ||| |   ||   |   |   |  ||   || |||     ||| ||| |
Db        945 TGTGTGTGCTTGCAGCTGTGTGTGCTTGCAGGTGCGTGTGCTTGCAGCTGTGCGTGCTTG 1004

Qy        270 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 329
              |     ||| |||| || |   ||  |||||||  |   |  |||||||  |  ||   |
Db       1005 CAGCTGTGCGTGCTTGCAGCTGTGTGTGCTTGCAGGTGTGTGTGCTTGCAGGTGCGTGTG 1064

Qy        330 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 389
              | |||     |||  ||| ||     ||| |||| || |   ||  |||||||       |
Db       1065 CTTGCAGCTGTGCGTGCTTGCAGCTGTGCGTGCTTGCAGCTGTGTGTGCTTGCAGCTGTG 1124

Qy        390 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAA 449
                |||||||| |  ||   || |||  |     | |  || ||| ||  |||||||||
Db       1125 TGTGCTTGCTGGTGCGTGTGCTTGCAGGTGTGTGTGCTAGCAGGTGTGTGTGCTTGCTGG 1184

Qy        450 GGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 509
              | |   |   ||  |   |  |||||   ||  |  |||| |   |   ||  |  ||
Db       1185 TGCGTGTGCTTGCTGGTGCGTGTGCTTGCAGGTGCGTGTGCTTGCAGGTGTGTGTGCTTG 1244

Qy        510 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTC 569
              ||||   |   |        |   |  |||| | |||  ||   |  |||| | |||  |
Db       1245 CTGGTGCGTGTGCTTACGGGTGTGTGTGCTTGCAGGTGTGTGTG--TGCTTGCAGGTGTG 1302

Qy        570 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTG 629
                |  |||| ||  |  || | |   |   |  |    ||||   |  ||  |  |
Db       1303 TGTGTTTGCTGGTGCGTGTGTTTGCAGGTGCATGTGCTAGCTGGTGCGTGTGTTTGCAGG 1362

Qy        630 CTGGGCTGCTAGTAGGGCTGCTGGGCTG 657
                |  |||| | |||  ||   | |||
Db       1363 TGCGTGTGCTTGCAGGTGTGTGTGACTG 1390


RESULT 5
```

```
US-10-027-632-164203/c
; Sequence 164203, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 164203
;   LENGTH: 798
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-164203

  Query Match             9.2%;  Score 92.2;  DB 13;  Length 798;
  Best Local Similarity   45.9%;  Pred. No. 3.1e-16;
  Matches  307;  Conservative    2;  Mismatches  360;  Indels    0;  Gaps    0;

Qy         67 ACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGC 126
              | |  ||   |||||   |   | |||  ||     |  |    |     |||  |||
Db        683 ATGCATGAGAGCCTGGATGAAAGACTGACTGGAAGGCTGTTTGACAAGCTGGCTGCCTAA 624

Qy        127 TGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGC 186
               |||||  |    ||  |   ||  |    ||| ||| || ||   |  |    ||| ||
Db        623 AAGCCTGGCTCAAAAGCTGCTGGTTGAGAGGCTGGCTTAGAGGCTTGCTGAGAAGCTGGC 564

Qy        187 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGC 246
              |   |     ||||||  || | |  || |    |||||||  |||   |||
Db        563 TGTATTAAAGGCTTGCCTAGAGCATGGTTGTCAGGCTGGCTGGCTGTCAGGCTGAGCAAG 504

Qy        247 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 306
              |     || |   ||| |      |||   || | | |||| |  |||     ||| |
Db        503 GGTTTATCTGGAAGGCTGGTCAAGATGCAGTCTGGGTGGCTGCCTGGCTGAGAGGCTGGT 444

Qy        307 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 366
              || |  |   ||| |||    |||   ||  ||      ||||| |    | || ||
Db        443 TGCCTGGGAAGCTGGCTGAGAGGCTGCCTGCCTGCAAGTCTGGCTTAGAGGGTTACTGAG 384

Qy        367 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 426
              |||  ||| | ||    ||| |||||   ||  ||| |  ||   ||| ||||||  ||
Db        383 AGGCTGGCTGGATTAGAGGCTTGCCTGGAGGCCAGCTAGTAGGTGGGCTGGCTGAGAGGC 324

Qy        427 CAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGG 486
              ||  |         |   | | ||  ||        ||  | |||   | |  | | |
Db        323 TAGCTGCTTGAAAGGGCAGTTGAGAGGCTGGCTGAGAGGCTGGCTGGGAGGCTGTTTGAG 264

Qy        487 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 546
              |||||  : ||   ||   |||  |  |||    ||     ||   |||   ||:
Db        263 AGGCTGGCTKCCTGAGAGACTGGCTCAGAGGCTGCCTGGTTGAGAGGCTGGCTTAGAGRC 204

Qy        547 CTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 606
              |  ||| |   || |  ||       |  |||||  ||    |||||   | | |  ||
Db        203 TTCCTGAGAGCCTGGCTGGTTGAGAAGCTTGCTAAGAGACTGGCTGGCAGGTTGGCTGGC 144

Qy        607 CTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 666
                  | |||| |   |||  ||| ||   || |    |  |  ||   |||    ||
Db        143 TGAGAGATCTGCCAAGGGGGTGGCTTGGAATCTTGCCCAGAGGTGGGCTGGCTCCCTGGT 84

Qy        667 CTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGC 726
              |   ||    || | |||    ||  | | ||  || |   ||| ||      |  || |
Db         83 TTAGAGGTTACCTCGAAGGCTCACCAACAGACTGCCTGGGAGGCTGGCAGAGAGACCCAC 24

Qy        727 TTGCCTGCT 735
                ||  |||
Db         23 AGGCTAGCT 15


RESULT 6
US-10-027-632-164204/c
; Sequence 164204, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 164204
;   LENGTH: 798
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-164204

  Query Match             9.2%;  Score 92.2;  DB 13;  Length 798;
  Best Local Similarity   45.9%;  Pred. No. 3.1e-16;
  Matches  307;  Conservative    2;  Mismatches  360;  Indels    0;  Gaps    0;

Qy         67 ACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGC 126
              | |  ||   |||||   |   | |||  ||     |  |    |     |||  |||
Db        683 ATGCATGAGAGCCTGGATGAAAGACTGACTGGAAGGCTGTTTGACAAGCTGGCTGCCTAA 624

Qy        127 TGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGC 186
               |||||  |    ||  |   ||  |    ||| ||| || ||   |  |    ||| ||
Db        623 AAGCCTGGCTCAAAAGCTGCTGGTTGAGAGGCTGGCTTAGAGGCTTGCTGAGAAGCTGGC 564

Qy        187 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGC 246
              |   |     ||||||  || | |  || |    |||||||  |||   |||
Db        563 TGTATTAAAGGCTTGCCTAGAGCATGGTTGTCAGGCTGGCTGGCTGTCAGGCTGAGCAAG 504

Qy        247 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 306
              |     || |   ||| |      |||   || | | |||| |  |||     ||| |
Db        503 GGTTTATCTGGAAGGCTGGTCAAGATGCAGTCTGGGTGGCTGCCTGGCTGAGAGGCTGGT 444

Qy        307 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 366
              || |  |   ||| |||    |||   ||  ||      ||||| |    | || ||
Db        443 TGCCTGGGAAGCTGGCTGAGAGGCTGCCTGCCTGCAAGTCTGGCTTAGAGGGTTACTGAG 384

Qy        367 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 426
              |||  ||| | ||    ||| |||||   ||  ||| |  ||   ||| ||||||  ||
Db        383 AGGCTGGCTGGATTAGAGGCTTGCCTGGAGGCCAGCTAGTAGGTGGGCTGGCTGAGAGGC 324

Qy        427 CAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGG 486
              ||  |         |   | | ||  ||        ||  | |||   | |  | | |
```

```
Db        143 TGAGAGATCTGCCAAGGGGGTGGCTTGGAATCTTGCCCAGAGGTGGGCTGGCTCCCTGGT 84

Qy        667 CTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGC 726
              |   ||    || | |||    ||  | | ||  || |   ||| ||      |  || |
Db         83 TTAGAGGTTACCTCGAAGGCTCACCAACAGACTGCCTGGGAGGCTGGCAGAGAGACCCAC 24

Qy        727 TTGCCTGCT 735
                ||  |||
Db         23 AGGCTAGCT 15


RESULT 6
US-10-027-632-164204/c
; Sequence 164204, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 164204
;   LENGTH: 798
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-164204

  Query Match             9.2%;  Score 92.2;  DB 13;  Length 798;
  Best Local Similarity   45.9%;  Pred. No. 3.1e-16;
  Matches  307;  Conservative    2;  Mismatches  360;  Indels    0;  Gaps    0;

Qy         67 ACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGC 126
              | |  ||   |||||   |   | |||  ||     |  |    |     |||  |||
Db        683 ATGCATGAGAGCCTGGATGAAAGACTGACTGGAAGGCTGTTTGACAAGCTGGCTGCCTAA 624

Qy        127 TGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGC 186
               |||||  |    ||  |   ||  |    ||| ||| || ||   |  |    ||| ||
Db        623 AAGCCTGGCTCAAAAGCTGCTGGTTGAGAGGCTGGCTTAGAGGCTTGCTGAGAAGCTGGC 564

Qy        187 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGC 246
              |   |     ||||||  || | |  || |    |||||||  |||   |||
Db        563 TGTATTAAAGGCTTGCCTAGAGCATGGTTGTCAGGCTGGCTGGCTGTCAGGCTGAGCAAG 504

Qy        247 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 306
              |     || |   ||| |      |||   || | | |||| |  |||     ||| |
Db        503 GGTTTATCTGGAAGGCTGGTCAAGATGCAGTCTGGGTGGCTGCCTGGCTGAGAGGCTGGT 444

Qy        307 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 366
              || |  |   ||| |||    |||   ||  ||      ||||| |    | || ||
Db        443 TGCCTGGGAAGCTGGCTGAGAGGCTGCCTGCCTGCAAGTCTGGCTTAGAGGGTTACTGAG 384

Qy        367 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 426
              |||  ||| | ||    ||| |||||   ||  ||| |  ||   ||| ||||||  ||
Db        383 AGGCTGGCTGGATTAGAGGCTTGCCTGGAGGCCAGCTAGTAGGTGGGCTGGCTGAGAGGC 324

Qy        427 CAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGG 486
              ||  |         |   | | ||  ||        ||  | |||   | |  | | |
```

```
Db        323 TAGCTGCTTGAAAGGGCAGTTGAGAGGCTGGCTGAGAGGCTGGCTGGGAGGCTGTTTGAG 264

Qy        487 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 546

Db        263 AGGCTGGCTKCCTGAGAGACTGGCTCAGAGGCTGCCTGGTTGAGAGGCTGGCTTAGAGRC 204

Qy        547 CTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 606

Db        203 TTCCTGAGAGCCTGGCTGGTTGAGAAGCTTGCTAAGAGACTGGCTGGCAGGTTGGCTGGC 144

Qy        607 CTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 666

Db        143 TGAGAGATCTGCCAAGGGGGTGGCTTGGAATCTTGCCCAGAGGTGGGCTGGCTCCCTGGT 84

Qy        667 CTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGC 726

Db         83 TTAGAGGTTACCTCGAAGGCTCACCAACAGACTGCCTGGGAGGCTGGCAGAGAGACCCAC 24

Qy        727 TTGCCTGCT 735

Db         23 AGGCTAGCT 15
```

RESULT 7
US-10-017-161-1915
; Sequence 1915, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1915
; LENGTH: 14750
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(14750)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(343)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5163)..(5441)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (13911)..(14088)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (14443)..(14550)
US-10-017-161-1915

Query Match 8.4%; Score 83.6; DB 12; Length 14750;
Best Local Similarity 46.5%; Pred. No. 1.2e-13;
Matches 336; Conservative 0; Mismatches 384; Indels 2; Gaps 2;

```
Qy         15 GAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGC 74

Db        896 GCGCTGAGATGCTGGGATGCTAAGGTGCTAAGGTGCTGGGGTGCTGGGGTGCTAGGATGC 955

Qy         75 CTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGC 134

Db        956 TGAGGTTTTGTGATACTGGGTCGCTGAGATGCTGGGATACTGGGGTGCTGAAGTGCTGGG 1015

Qy        135 TTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGC 194

Db       1016 GTACTGGGGTGTTGCCATGCTGAAATGCTGGGGTGTTGGGATGCTGAGGTGCTGGGTTGC 1075

Qy        195 CTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGC 254

Db       1076 TGGGATGCTG-GGATAGTCCTTGGATGCTGCGGTGCTGAGATGCTGGTCTGCTGGTCTGC 1134

Qy        255 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 314

Db       1135 TGGGGTGTTGGGATGCTGGGATTTTGGGATGCTGGAGTGCTGCAGTGCTGCAGTGCTGAG 1194

Qy        315 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 374

Db       1195 ATGCTGGAGTGCTGGGGTGCTGGAATACTGTAGTACTGGTGTGCTGGAGCATTGAGATGC 1254

Qy        375 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGG 434

Db       1255 TAGGGCACTGGGATGCTAAGGTGCTGAGATGCTGCAGTGCTGGGGTGTTGGGATGCCGAG 1314

Qy        435 CTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGG 494

Db       1315 GTGTTGGGATGCTTAAGTGCTGGGGTGTTGGGATGCTGGGGTGCTGGAACACTATGGTGC 1374

Qy        495 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 554

Db       1375 TGGGGTGCTGGAGTGTTGAGATACTGTAGTGGTGGGATGCTTAAGTGCTGGGGTGCTGGG 1434

Qy        555 CTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGG 614

Db       1435 -TGTTGGGATGCCTAGGTGCTGGGGTGCCGAGATGCTGGAGTACTAGTGTGCTGGGATGC 1493

Qy        615 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 674

Db       1494 TGAAGTGCTGGGGTCCTGAGATGCTGCAGTGCTAGGGTGCTGAGATTCTGGGCTGCTGGA 1553

Qy        675 CTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGC 734

Db       1554 GTGTTGGGGTGCTGGGATGCTGGAGTGCTGAGATGCTTGGACAATGGGGTGCTGGAATAC 1613

Qy        735 TA 736

Db       1614 TA 1615
```

RESULT 8
US-10-027-632-272985/c
; Sequence 272985, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 272985
; LENGTH: 572

```
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-272985

  Query Match             8.3%;  Score 82.6;  DB 13;  Length 572;
  Best Local Similarity   46.7%;  Pred. No. 1.4e-13;
  Matches  259;  Conservative    1;  Mismatches  295;  Indels    0;  Gaps    0;

Qy        181 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCT 240
              ||| |||   |     ||||||  || | |  || |    ||||||| |||   |||
Db        568 GCTGGCTGTATTAAAGGCTTGCCTAGAGCATGGTTGTCAGGCTGGCTGGCTGTCAGGCTG 509

Qy        241 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 300
                  |     || |  ||| |      |||   || | | |||| |  |||
Db        508 AGCAAGGGTTTATCTGGAAGGCTGGTCAAGATGCAGTCTGGGTGGCTGCCTGGCTGAGAG 449

Qy        301 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 360
              ||| | || |  |   ||| |||    |||   ||  ||      |||||  |    | ||
Db        448 GCTGGTTGCCTGGGAAGCTGGCTGAGAGGCTGCCTGCCTGCAAGTCTGGCTTAGAGGGTT 389

Qy        361 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 420
               ||    |||  ||| | ||    ||| |||||   ||  :|| |  ||   ||| ||||
Db        388 ACTGAGAGGCTGGCTGGATTAGAGGCTTGCCTGGAGGCCAKCTAGTAGGTGGGCTGGCTG 329

Qy        421 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 480
              || ||  ||  |         |   | | || ||  ||        ||  | |||   | |
Db        328 AGAGGCTAGCTGCTTGAAAGGGCAGTTGAGAGGCTGGCTGAGAGGCTGGCTGGGAGGCTG 269

Qy        481 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 540
               | | |  |||||  | ||   ||    |||  |  |||    ||      ||   |||
Db        268 TTTGAGAGGCTGGCTTCCTGAGAGACTGGCTCAGAGGCTGCCTGGTTGAGAGGCTGGCTT 209

Qy        541 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 600
              |||  | |||  |   || |  ||      |  |||||  ||    |||||   | |
Db        208 AGAGGCTTCCTGAGAGCCTGGCTGGTTGAGAAGCTTGCTAAGAGACTGGCTGGCAGGTTG 149

Qy        601 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 660
              |  ||      |  |||| |   |||  ||| ||  ||  |    |  |  ||   |||
Db        148 GCTGGCTGAGAGATCTGCCAAGGGGGTGGCTTGGAATCTTGCCCAGAGGTGGGCTGGCTC 89

Qy        661 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGG 720
                 ||  |   ||    || | |||    ||  | | ||  || |   ||| ||     |
Db         88 CCTGGTTTAGAGGTTACCTCGAAGGCTCACCAACAGACTGCCTGGGAGGCTGGCAGAGAG 29

Qy        721 GGCCGCTTGCCTGCT 735
                ||  |  ||  |||
Db         28 ACCCACAGGCTAGCT 14


RESULT 9
US-09-799-462A-17
; Sequence 17, Application US/09799462A
; Patent No. US20020160970A1
;   GENERAL INFORMATION:
;        APPLICANT: Hadlaczky, Gyula
;                   Szalay, Aladar
;        TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
;                            AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
;        NUMBER OF SEQUENCES: 34
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Heller Ehrman White & McAuliffe
;             STREET: 4250 Executive Square, 7th Floor
;             CITY: La Jolla
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 92037
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ Version 1.5
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/799,462A
;             FILING DATE: 10-Sep-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/835,682
;             FILING DATE: 10-APR-1997
;             APPLICATION NUMBER: 08/695,191
;             FILING DATE: 07-AUG-1996
;             APPLICATION NUMBER: 08/682,080
;             FILING DATE: 15-JUL-1996
;             APPLICATION NUMBER: 08/629,822
;             FILING DATE: 10-APR-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Seidman, Stephanie L
;             REGISTRATION NUMBER: 33,779
;             REFERENCE/DOCKET NUMBER: 24601-402G
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 858-450-8403
;             TELEFAX: 858-587-5360
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 17:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 42999 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: Genomic DNA
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        FRAGMENT TYPE: <Unknown>
;        ORIGINAL SOURCE:
;        SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-799-462A-17

  Query Match             8.1%;  Score 80.8;  DB 10;  Length 42999;
  Best Local Similarity   58.1%;  Pred. No. 8.3e-13;
  Matches  179;  Conservative    0;  Mismatches  127;  Indels    2;  Gaps    2;

Qy        114 CCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAG 173
              | | |||||| |||  | ||||||||         |  | | || ||   ||
Db      40275 CTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTGTCTTGTTTCTCGATTTCTTTC 40334

Qy        174 TAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAG 233
              |     |  || ||  |  || ||||| ||     ||  |      ||||  ||||||
Db      40335 TTTCTTTTGTTTCTTTCCTGCTTGCTTTCTTGCTTGATTGCTTTCGTGCTTTCTTGCTTT 40394

Qy        234 TAGGGCTGCTCGCTGGC-CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 292
                |  | ||  ||  |  |  || ||| ||  || || | |||| || || |||  | |
Db      40395 CTTGTTTTCTTTCTTTCTTTTGTTTCTTTCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTT 40454

Qy        293 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC-TTGCTTGCTCGCTGGCC 351
              ||||||||||||  |  || | |||| || || |||  | | | ||  || ||  || ||
Db      40455 GCTTGCTTGCTTTCGTGCTTTCTTGTTTTCTTGCTTTCTTTCTTTTGTTTCTTTCTTGCT 40514

Qy        352 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              ||||| ||||||  || |  | ||||||| || ||| || ||||| | ||||  ||  |
Db      40515 TGCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTTCTTTCT 40574

Qy        412 TGCTTGCT 419
              |||||  ||
Db      40575 TGCTTTCT 40582


RESULT 10
US-09-836-911A-17
; Sequence 17, Application US/09836911A
; Publication No. US20030033617A1
;   GENERAL INFORMATION:
;        APPLICANT: Hadlaczky, Gyula
;                   Szalay, Aladar
;        TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
;                            AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
;        NUMBER OF SEQUENCES: 34
;        CORRESPONDENCE ADDRESS:
```

```
;           ADDRESSEE: Heller Ehrman White & McAuliffe
;           STREET: 4350 La Jolla Village Drive, 6th Floor
;           CITY: San Diego
;           STATE: CA
;           COUNTRY: USA
;           ZIP: 92122
;         COMPUTER READABLE FORM:
;           MEDIUM TYPE: Diskette
;           COMPUTER: IBM Compatible
;           OPERATING SYSTEM: DOS
;           SOFTWARE: FastSEQ Version 1.5
;         CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/09/836,911A
;           FILING DATE: 17-Apr-2002
;           CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: 08/835,682
;           FILING DATE: 10-APR-1997
;           APPLICATION NUMBER: 08/695,191
;           FILING DATE: 07-AUG-1996
;           APPLICATION NUMBER: 08/682,080
;           FILING DATE: 15-JUL-1996
;           APPLICATION NUMBER: 08/629,822
;           FILING DATE: 10-APR-1996
;         ATTORNEY/AGENT INFORMATION:
;           NAME: Seidman, Stephanie L
;           REGISTRATION NUMBER: 33,779
;           REFERENCE/DOCKET NUMBER: 24601-402I
;         TELECOMMUNICATION INFORMATION:
;           TELEPHONE: 858-450-8403
;           TELEFAX: 858-587-5360
;           TELEX: <Unknown>
;    INFORMATION FOR SEQ ID NO: 17:
;         SEQUENCE CHARACTERISTICS:
;           LENGTH: 42999 base pairs
;           TYPE: nucleic acid
;           STRANDEDNESS: single
;           TOPOLOGY: linear
;         MOLECULE TYPE: Genomic DNA
;         HYPOTHETICAL: NO
;         ANTI-SENSE: NO
;         FRAGMENT TYPE: <Unknown>
;         ORIGINAL SOURCE:
;         SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-836-911A-17

  Query Match             8.1%;  Score 80.8;  DB 11;  Length 42999;
  Best Local Similarity   58.1%;  Pred. No. 8.3e-13;
  Matches  179;  Conservative    0;  Mismatches  127;  Indels    2;  Gaps    2;

Qy        114 CCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAG 173
              | |*|||||| |||  | ||||||||         |  | | || ||   ||
Db      40275 CTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTGTCTTGTTTCTCGATTTCTTTC 40334

Qy        174 TAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAG 233
              |     |  || ||  |  || ||||| ||     ||  |      |||| ||||||
Db      40335 TTTCTTTTGTTTCTTTCCTGCTTGCTTTCTTGCTTGATTGCTTTCGTGCTTTCTTGCTTT 40394

Qy        234 TAGGGCTGCTCGCTGGC-CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 292
                | | ||  ||  |  |  || ||| ||  || || | |||| || || |||  | |
Db      40395 CTTGTTTTCTTTCTTTCTTTTGTTTCTTTCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTT 40454

Qy        293 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC-TTGCTTGCTCGCTGGCC 351
              ||||||||||  |  || | |||| || || |||  | | |  ||  || ||  || ||
Db      40455 GCTTGCTTGCTTTCGTGCTTTCTTGTTTTCTTGCTTTCTTTCTTTTGTTTCTTTCTTGCT 40514

Qy        352 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              ||||| ||||||  || |  | ||||||| || ||| || ||||| | ||||  ||  |
Db      40515 TGCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTTCTTTCT 40574

Qy        412 TGCTTGCT 419
              ||||| ||
Db      40575 TGCTTTCT 40582
```

```
RESULT 11
US-10-125-767-17
; Sequence 17, Application US/10125767
; Publication No. US20020160410A1
;   GENERAL INFORMATION:
;         APPLICANT: Hadlaczky, Gyula
;         TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND
;                             METHODS
;                             FOR PREPARING ARTIFICIAL CHROMOSOMES
;         NUMBER OF SEQUENCES: 34
;         CORRESPONDENCE ADDRESS:
;           ADDRESSEE: Heller Ehrman White & McAuliffe LLP
;           STREET: 4350 La Jolla Village Drive, 7th Floor
;           CITY: San Diego
;           STATE: CA
;           COUNTRY: USA
;           ZIP: 92122
;         COMPUTER READABLE FORM:
;           MEDIUM TYPE: Diskette
;           COMPUTER: IBM Compatible
;           OPERATING SYSTEM: DOS
;           SOFTWARE: FastSEQ Version 1.5
;         CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/10/125,767
;           FILING DATE: 17-Apr-2002
;         PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: 09/724,693
;           FILING DATE: 28-NOV-2000
;           APPLICATION NUMBER: 08/835,682
;           FILING DATE: 10-APR-1997
;           APPLICATION NUMBER: 08/695,191
;           FILING DATE: 07-AUG-1996
;           APPLICATION NUMBER: 08/682,080
;           FILING DATE: 15-JUL-1996
;           APPLICATION NUMBER: 08/629,822
;           FILING DATE: 10-APR-1996
;         ATTORNEY/AGENT INFORMATION:
;           NAME: Seidman, Stephanie L
;           REGISTRATION NUMBER: 33,779
;           REFERENCE/DOCKET NUMBER: 24601-402J
;         TELECOMMUNICATION INFORMATION:
;           TELEPHONE: 858-450-8403
;           TELEFAX: 858-587-5360
;           TELEX: <Unknown>
;    INFORMATION FOR SEQ ID NO: 17:
;         SEQUENCE CHARACTERISTICS:
;           LENGTH: 42999 base pairs
;           TYPE: nucleic acid
;           STRANDEDNESS: single
;           TOPOLOGY: linear
;         MOLECULE TYPE: Genomic DNA
;         HYPOTHETICAL: NO
;         ANTI-SENSE: NO
;         FRAGMENT TYPE: <Unknown>
;         ORIGINAL SOURCE:
;         SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-125-767-17

  Query Match             8.1%;  Score 80.8;  DB 13;  Length 42999;
  Best Local Similarity   58.1%;  Pred. No. 8.3e-13;
  Matches  179;  Conservative    0;  Mismatches  127;  Indels    2;  Gaps    2;

Qy        114 CCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAG 173
              | | |||||| |||  | ||||||||         |  | | || ||   ||
Db      40275 CTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTGTCTTGTTTCTCGATTTCTTTC 40334

Qy        174 TAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAG 233
              |     |  || ||  |  || ||||| ||     ||  |      |||| ||||||
Db      40335 TTTCTTTTGTTTCTTTCCTGCTTGCTTTCTTGCTTGATTGCTTTCGTGCTTTCTTGCTTT 40394

Qy        234 TAGGGCTGCTCGCTGGC-CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 292
```

```
Db      40395 CTTGTTTTCTTTCTTTCTTTTGTTTCTTTCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTT 40454

Qy        293 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC-TTGCTTGCTCGCTGGCC 351
              ||||||||||  |  || | |||| || || |||  | | | ||  || ||  || ||
Db      40455 GCTTGCTTGCTTTCGTGCTTTCTTGTTTTCTTGCTTTCTTTCTTTTGTTTCTTTCTTGCT 40514

Qy        352 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              ||||| ||||||  || |  | ||||||| || ||| || ||||| | ||||  ||  |
Db      40515 TGCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTTCTTTCT 40574

Qy        412 TGCTTGCT 419
              ||||| ||
Db      40575 TGCTTTCT 40582


RESULT 12
US-10-151-081-17
; Sequence 17, Application US/10151081
; Publication No. US20030083293A1
;   GENERAL INFORMATION:
;        APPLICANT: Hadlaczky, Gyula
;                   Szalay, Aladar
;        TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
;                            AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
;        NUMBER OF SEQUENCES: 34
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Heller Ehrman White & McAuliffe
;             STREET: 4250 Executive Square, 7th Floor
;             CITY: La Jolla
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 92037
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ Version 1.5
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/151,081
;             FILING DATE: 16-May-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/799,462
;             FILING DATE: HEREWITH  05-MAR-2001
;             APPLICATION NUMBER: 09/724,693
;             FILING DATE: HEREWITH  28-NOV-2000
;             APPLICATION NUMBER: 08/835,682
;             FILING DATE: 10-APR-1997
;             APPLICATION NUMBER: 08/695,191
;             FILING DATE: 07-AUG-1996
;             APPLICATION NUMBER: 08/682,080
;             FILING DATE: 15-JUL-1996
;             APPLICATION NUMBER: 08/629,822
;             FILING DATE: 10-APR-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Seidman, Stephanie L
;             REGISTRATION NUMBER: 33,779
;             REFERENCE/DOCKET NUMBER: 24601-402L
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 858-450-8403
;             TELEFAX: 858-587-5360
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 17:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 42999 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: Genomic DNA
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        FRAGMENT TYPE: <Unknown>
;        ORIGINAL SOURCE:
;        SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-151-081-17

  Query Match             8.1%;  Score 80.8;  DB 14;  Length 42999;
  Best Local Similarity   58.1%;  Pred. No. 8.3e-13;
  Matches  179;  Conservative    0;  Mismatches  127;  Indels    2;  Gaps    2;

Qy        114 CCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAG 173
              | | |||||| |||  | ||||||||        |  | | || ||   ||
Db      40275 CTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTGTCTTGTTTCTCGATTTCTTTC 40334

Qy        174 TAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAG 233
              |      |  || ||  |  || ||||| ||      ||  |      ||||  ||||||
Db      40335 TTTCTTTTGTTTCTTTCCTGCTTGCTTTCTTGCTTGATTGCTTTCGTGCTTTCTTGCTTT 40394

Qy        234 TAGGGCTGCTCGCTGGC-CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 292
                | | ||  ||  |  |  || ||| ||  || || | |||| || || |||  | |
Db      40395 CTTGTTTTCTTTCTTTCTTTTGTTTCTTTCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTT 40454

Qy        293 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC-TTGCTTGCTCGCTGGCC 351
              ||||||||||  |  || | |||| || || |||  | | | ||  || ||  || ||
Db      40455 GCTTGCTTGCTTTCGTGCTTTCTTGTTTTCTTGCTTTCTTTCTTTTGTTTCTTTCTTGCT 40514

Qy        352 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              ||||| ||||||  || |  | ||||||| || ||| || ||||| | ||||  ||  |
Db      40515 TGCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTTCTTTCT 40574

Qy        412 TGCTTGCT 419
              ||||| ||
Db      40575 TGCTTTCT 40582


RESULT 13
US-10-287-313-17
; Sequence 17, Application US/10287313
; Publication No. US20030101480A1
;   GENERAL INFORMATION:
;        APPLICANT: Hadlaczky, Gyula
;                   Szalay, Aladar
;        TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
;                            AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
;        NUMBER OF SEQUENCES: 34
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Heller Ehrman White & McAuliffe
;             STREET: 4250 Executive Square, 7th Floor
;             CITY: La Jolla
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 92037
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ Version 1.5
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/287,313
;             FILING DATE: 01-No. US20030101480A1-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/724,726
;             FILING DATE: 28-NOV-2000
;             APPLICATION NUMBER: 08/835,682
;             FILING DATE: 10-APR-1997
;             APPLICATION NUMBER: 08/695,191
;             FILING DATE: 07-AUG-1996
;             APPLICATION NUMBER: 08/682,080
;             FILING DATE: 15-JUL-1996
;             APPLICATION NUMBER: 08/629,822
;             FILING DATE: 10-APR-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Seidman, Stephanie L
;             REGISTRATION NUMBER: 33,779
```

```
;           REFERENCE/DOCKET NUMBER: 6869-402N
;       TELECOMMUNICATION INFORMATION:
;           TELEPHONE: 858-450-8403
;           TELEFAX: 858-587-5360
;           TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 17:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH: 42999 base pairs
;           TYPE: nucleic acid
;           STRANDEDNESS: single
;           TOPOLOGY: linear
;       MOLECULE TYPE: Genomic DNA
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       FRAGMENT TYPE: <Unknown>
;       ORIGINAL SOURCE:
;       SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-287-313-17

  Query Match             8.1%;  Score 80.8;  DB 14;  Length 42999;
  Best Local Similarity   58.1%;  Pred. No. 8.3e-13;
  Matches  179;  Conservative    0;  Mismatches  127;  Indels    2;  Gaps    2;

Qy        114 CCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAG 173
              | | |||||| |||  | ||||||||         |  | | || ||   ||
Db      40275 CTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTGTCTTGTTTCTCGATTTCTTTC 40334

Qy        174 TAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAG 233
              |     |  || ||  |  || ||||| ||     ||  |      |||| ||||||
Db      40335 TTTCTTTTGTTTCTTTCCTGCTTGCTTTCTTGCTTGATTGCTTTCGTGCTTTCTTGCTTT 40394

Qy        234 TAGGGCTGCTCGCTGGC-CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 292
                | | ||- ||  |  |  || ||| ||  || || | |||| || || |||  | |
Db      40395 CTTGTTTTCTTTCTTTCTTTTGTTTCTTTCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTT 40454

Qy        293 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC-TTGCTTGCTCGCTGGCC 351
              ||||||||||||  |  || | |||| || || |||  | | | ||  || ||  || ||
Db      40455 GCTTGCTTGCTTTCGTGCTTTCTTGTTTTCTTGCTTTCTTTCTTTTGTTTCTTTCTTGCT 40514

Qy        352 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              |||||  ||||||  || |  | ||||||| || ||| || |||| | ||||  || |
Db      40515 TGCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTTCTTTCT 40574

Qy        412 TGCTTGCT 419
              |||||  ||
Db      40575 TGCTTTCT 40582


RESULT 14
US-10-219-694-17
; Sequence 17, Application US/10219694
; Publication No. US20030108914A1
;   GENERAL INFORMATION:
;       APPLICANT: Hadlaczky, Gyula
;       TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS
;                           FOR PREPARING ARTIFICIAL CHROMOSOMES
;       NUMBER OF SEQUENCES: 34
;       CORRESPONDENCE ADDRESS:
;           ADDRESSEE: Heller Ehrman White & McAuliffe LLP
;           STREET: 4350 La Jolla Village Drive, 7th Floor
;           CITY: San Diego
;           STATE: CA
;           COUNTRY: USA
;           ZIP: 92122
;       COMPUTER READABLE FORM:
;           MEDIUM TYPE: Diskette
;           COMPUTER: IBM Compatible
;           OPERATING SYSTEM: DOS
;           SOFTWARE: FastSEQ Version 1.5
;       CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/10/219,694
;           FILING DATE: 14-Aug-2002
;           CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: <Unknown>
;           FILING DATE: 28-NOV-2000
;           APPLICATION NUMBER: 08/835,682
;           FILING DATE: 10-APR-1997
;           APPLICATION NUMBER: 08/695,191
;           FILING DATE: 07-AUG-1996
;           APPLICATION NUMBER: 08/682,080
;           FILING DATE: 15-JUL-1996
;           APPLICATION NUMBER: 08/629,822
;           FILING DATE: 10-APR-1996
;       ATTORNEY/AGENT INFORMATION:
;           NAME: Seidman, Stephanie L
;           REGISTRATION NUMBER: 33,779
;           REFERENCE/DOCKET NUMBER: 24601-402M
;       TELECOMMUNICATION INFORMATION:
;           TELEPHONE: 858-450-8403
;           TELEFAX: 858-587-5360
;           TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 17:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH: 42999 base pairs
;           TYPE: nucleic acid
;           STRANDEDNESS: single
;           TOPOLOGY: linear
;       MOLECULE TYPE: Genomic DNA
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       FRAGMENT TYPE: <Unknown>
;       ORIGINAL SOURCE:
;       SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-219-694-17

  Query Match             8.1%;  Score 80.8;  DB 14;  Length 42999;
  Best Local Similarity   58.1%;  Pred. No. 8.3e-13;
  Matches  179;  Conservative    0;  Mismatches  127;  Indels    2;  Gaps    2;

Qy        114 CCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAG 173
              | | |||||| |||  | ||||||||         |  | | || ||   ||
Db      40275 CTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTGTCTTGTTTCTCGATTTCTTTC 40334

Qy        174 TAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAG 233
              |     |  || ||  |  || ||||| ||     ||  |      |||| ||||||
Db      40335 TTTCTTTTGTTTCTTTCCTGCTTGCTTTCTTGCTTGATTGCTTTCGTGCTTTCTTGCTTT 40394

Qy        234 TAGGGCTGCTCGCTGGC-CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 292
                | | ||- ||  |  |  || ||| ||  || || | |||| || || |||  | |
Db      40395 CTTGTTTTCTTTCTTTCTTTTGTTTCTTTCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTT 40454

Qy        293 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC-TTGCTTGCTCGCTGGCC 351
              ||||||||||||  |  || | |||| || || |||  | | | ||  || ||  || ||
Db      40455 GCTTGCTTGCTTTCGTGCTTTCTTGTTTTCTTGCTTTCTTTCTTTTGTTTCTTTCTTGCT 40514

Qy        352 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              |||||  ||||||  || |  | ||||||| || ||| || |||| | ||||  || |
Db      40515 TGCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTTCTTTCT 40574

Qy        412 TGCTTGCT 419
              |||||  ||
Db      40575 TGCTTTCT 40582


RESULT 15
US-10-017-161-2193
; Sequence 2193, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
```

```
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2193
;   LENGTH: 1297
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1297)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1097)
US-10-017-161-2193

  Query Match             8.0%;  Score 80;  DB 12;  Length 1297;
  Best Local Similarity   59.1%;  Pred. No. 8.4e-13;
  Matches  208;  Conservative    0;  Mismatches  140;  Indels    4;  Gaps    4;

Qy         72 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGC- 130
Db        581 TGCTTGCTTGCTTGCTTTCGTGCTTTCTTGCTTTCCTGTTTTCTTTCTTTCTTTCTTCT 640

Qy        131 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGC 190
Db        641 TTTGTTTCTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTTCTTGTTTTCTCGC 700

Qy        191 TGGCCTGCTTGCTGAGGGGACAGTAGGGC-TGCTGGCTTGCTAGTAGGGCTGCTCGCTGG 249
Db        701 TTTCTTTCTTTCTTTTGTTTCTTTCCTGCTTGCTTTCTTGCTTGATTGCTTTCGTGCTTT 760

Qy        250 CCTGCTTGCTTGCTCGCTGGCCTGC-TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 308
Db        761 CTTGCTTTCTTGTTTTCTTTCTTTCTTTTGTTTCTTTCTTTCTTGCTTCCTTGTTTTCTT 820

Qy        309 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC-TTGCTCGCT 367
Db        821 GCTTTCTTGCTTGCTTGCTTTCGTGCTTTCTTGCTTTCTTTGCTTTCTTTCTTTTGTTTCT 880

Qy        368 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
Db        881 TTCTTGCTTGCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCT 932


RESULT 16
US-10-027-632-113494
; Sequence 113494, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 113494
;   LENGTH: 3213
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-113494

  Query Match             7.9%;  Score 79.4;  DB 13;  Length 3213;
  Best Local Similarity   47.3%;  Pred. No. 1.4e-12;
  Matches  303;  Conservative    0;  Mismatches  336;  Indels    2;  Gaps    2;

Qy         81 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 140
Db       1947 GGTAGGTGGATGGATGGATGAGTAGGTGGATGGATGGATGGATAGGTGGGTGGATGGGTG 2006

Qy        141 AGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTT 200
Db       2007 GGTGAGTGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGGTAGGTGAATG 2066

Qy        201 GCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTT 260
Db       2067 GGTCAGTGGA-TGGATGGATGGATGATGGATAGGTGAATGGGTCAGTGGATGGATGGATG 2125

Qy        261 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 320
Db       2126 GGTAGGTGGATGGATCTATGGGTAGGTGGATGGATCTATGGGTAGGTGGATGGGTGGGTG 2185

Qy        321 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 380
Db       2186 GGTAGATGGATGGCCAGATGGGTGGATGACTGGGTAGATGGATAGGTGAGTGGATGGATG 2245

Qy        381 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTT 440
Db       2246 GATGGATGGATAGATGGATGGGTAGGTGGAC-GGTGGATGGATGAATGGGTAGGTGGATG 2304

Qy        441 GCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTA 500
Db       2305 GATGGATAAGTTAATGAATGGATGGGTAGGTGGACGGATGGACGGGTAGGTGGATGGATG 2364

Qy        501 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 560
Db       2365 GATAGGTGGGTGGATGGATGGGTGAGTGGATGGGTGGGTGAGTGGGTGGATGGATGGATG 2424

Qy        561 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTA 620
Db       2425 GATGGATGAATGGGTCGGTGGATGGACGGATGGGTAGGTGGATGGATCTATGGGTAGGTG 2484

Qy        621 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 680
Db       2485 GATGGGTGGGTGGGTGGATGGGTGGGTGGGTAGATGGATGGATGGATGGGTGGACGAGTG 2544

Qy        681 GTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGG 721
Db       2545 GGTAGGTAGATAGGTGAGTGAATGGATGGATGGATAGATGG 2585


RESULT 17
US-09-728-446-295/c
; Sequence 295, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 295
```

```
;   LENGTH: 388
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(388)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-295

  Query Match             7.8%;  Score 78.4;  DB 9;  Length 388;
  Best Local Similarity   60.2%;  Pred. No. 2e-12;
  Matches  130;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
              | ||||  | |  |     ||||   | ||||   |  | ||  | | ||| | | |
Db        218 CCTGCTGTTGGTCCCTGCTGCTGCATTTGCTGCTTTTGCTATTGCTGGATGGCTGGATGG 159

Qy        286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
              ||||  |||| | | ||| | ||||  | ||||| | | |||||   ||| | ||| | |
Db        158 CTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGGTGG 99

Qy        346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 405
               ||||  | | ||| | | |||||  |||| | | ||| | ||||  | ||| | | | |
Db         98 ATGGCTGGATGGCTGGATGGCTGGTTTGCTGGATGGCTGGATGGCTGGGTTGGTGGATGG 39

Qy        406 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTG 441
              ||||   ||| | |  | |||  || ||  |||| |
Db         38 CTGGATGGCTGGGTTGGTGGATTGTTGGATTGCTGG 3


RESULT 18
US-09-728-446-369/c
; Sequence 369, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 369
;   LENGTH: 428
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-728-446-369

  Query Match             7.8%;  Score 78.4;  DB 9;  Length 428;
  Best Local Similarity   60.2%;  Pred. No. 2e-12;
  Matches  130;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
              | ||||  | |  |     ||||   |  |||||   |  | ||  | | ||| | | |
Db        259 CCTGCTGTTGGTCCCTGCTGCTGCATTTGCTGCTTTTGCTATTGCTGGATGGCTGGATGG 200

Qy        286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
              ||||  |||| | | ||| | ||||  | ||||| | | |||||   ||| | ||| | |
Db        199 CTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGGTGG 140

Qy        346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 405
               ||||  | | ||| | | |||||  |||| | | ||| | ||||  | ||| | | | |
Db        139 ATGGCTGGATGGCTGGATGGCTGGTTTGCTGGATGGCTGGATGGCTGGGTTGGTGGATGG 80

Qy        406 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTG 441
              ||||   ||| | |  | |||  || ||  |||| |
Db         79 CTGGATGGCTGGGTTGGTGGATTGTTGGATTGCTGG 44


RESULT 19
US-09-728-446-354/c
; Sequence 354, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 354
;   LENGTH: 436
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(436)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-354

  Query Match             7.8%;  Score 78.4;  DB 9;  Length 436;
  Best Local Similarity   60.2%;  Pred. No. 2e-12;
  Matches  130;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
              | ||||  | |  |     ||||   |  |||||   |  | ||  | | ||| | | |
Db        267 CCTGCTGTTGGTCCCTGCTGCTGCATTTGCTGCTTTTGCTATTGCTGGATGGCTGGATGG 208

Qy        286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
              ||||  |||| | | ||| | ||||  | ||||| | | |||||   ||| | ||| | |
Db        207 CTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGGTGG 148

Qy        346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 405
               ||||  | | ||| | | |||||  |||| | | ||| | ||||  | ||| | | | |
Db        147 ATGGCTGGATGGCTGGATGGCTGGTTTGCTGGATGGCTGGATGGCTGGGTTGGTGGATGG 88

Qy        406 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTG 441
              ||||   ||| | |  | |||  || ||  |||| |
Db         87 CTGGATGGCTGGGTTGGTGGATTGTTGGATTGCTGG 52


RESULT 20
US-09-728-446-362/c
; Sequence 362, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 362
;   LENGTH: 409
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(409)
```

```
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-362

  Query Match             7.7%;  Score 77.4;  DB 9;  Length 409;
  Best Local Similarity   59.7%;  Pred. No. 3.8e-12;
  Matches  129;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy        226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
              | ||||  | |  |     ||||   |   |||||   |  | ||  | | ||| | | |
Db        246 CCTGCTGTTGGTCCCTGCTGCTGCATTTGCTGCTTTTGCTATTGCTGGATGGCTGGATGG 187

Qy        286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
              ||||  |||| | | ||| | ||||  | ||||| | | |||||   ||| | ||| | |
Db        186 CTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGGTGG 127

Qy        346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 405
              ||||  | | ||| | | |||||  |||| | | ||| | ||||  | ||| | | | |
Db        126 ATGGCTGGATGGCTGGATGGCTGGTTTGCTGGATGGCTGGATGGCTGGGTTGGTGGATGG 67

Qy        406 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTG 441
              ||||   ||| | |  | |||  || ||  || | |
Db         66 CTGGATGGCTGGGTTGGTGGATTGTTGGATTGNTGG 31


RESULT 21
US-10-329-960-1
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
; GENERAL INFORMATION:
;  APPLICANT: Fleischmann et al.
;  TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
;  TITLE OF INVENTION:  Thereof, and Uses Thereof
;  FILE REFERENCE: PB186P1
;  CURRENT APPLICATION NUMBER: US/10/329,960
;  CURRENT FILING DATE:  2003-01-02
;  PRIOR APPLICATION NUMBER: US 09/643,990
;  PRIOR FILING DATE: 2000-08-23
;  PRIOR APPLICATION NUMBER: US 08/487,429
;  PRIOR FILING DATE: 1995-06-07
;  PRIOR APPLICATION NUMBER: US 08/426,787
;  PRIOR FILING DATE: 1995-04-21
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1830121
;   TYPE: DNA
;   ORGANISM: Haemophilus influenzae
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4747)..(4747)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9921)..(9921)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (10150)..(10150)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (29298)..(29298)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (36543)..(36543)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (36551)..(36551)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (36636)..(36636)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (40808)..(40810)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (44416)..(44416)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (44905)..(44905)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (44975)..(44975)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (45593)..(45593)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (45732)..(45732)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (47036)..(47036)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51334)..(51334)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51602)..(51602)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51786)..(51786)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51805)..(51805)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (55369)..(55369)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (65309)..(65309)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (65313)..(65313)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (80024)..(80024)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (100091)..(100091)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (102696)..(102696)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (105121)..(105121)
```

```
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (107248)..(107248)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (117136)..(117136)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (119750)..(119750)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (119924)..(119924)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (120038)..(120038)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (121344)..(121344)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (122167)..(122167)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (122336)..(122336)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (131340)..(131340)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (131360)..(131360)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (139910)..(139910)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (140398)..(140398)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (142750)..(142750)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (145058)..(145058)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (145171)..(145171)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (145942)..(145942)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (147197)..(147197)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (150841)..(150841)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (152500)..(152500)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (152530)..(152530)

  Query Match             7.7%;  Score 77.4;  DB 14;  Length 1830121;
  Best Local Similarity   84.5%;  Pred. No. 1.3e-11;
  Matches   87;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy        319 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 378
              ||||| ||| || ||||||||||||| ||| || ||||||||||||| ||| || |||||||
Db    1543129 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 1543188

Qy        379 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGA 421
              ||||| ||| || |||||||||||| ||| || ||||||| ||
Db    1543189 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCAGA 1543231


RESULT 22
US-09-764-891-8904
; Sequence 8904, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC006
;  CURRENT APPLICATION NUMBER: US/09/764,891
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 10231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8904
;   LENGTH: 368
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-891-8904

  Query Match             7.7%;  Score 77.2;  DB 11;  Length 368;
  Best Local Similarity   64.1%;  Pred. No. 4.2e-12;
  Matches  132;  Conservative    0;  Mismatches   73;  Indels    1;  Gaps    1;

Qy        215 AGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 274
              | |  |||  |||||| |   |  |||  |  || | ||||||| | || |  | |||
Db         99 ATGCTTGCTTTCTTGCTTGCTTGCTTGCTTTCGTGCTTTCTTGCTTTCCCGTTTTCTTGC 158

Qy        275 TTGCTTGCTCGC-TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 333
              || ||| ||  | |    | ||| |||||| |||  | ||||||||||||  |  || |
Db        159 TTTCTTTCTTTCTTTTGTTTCTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTT 218

Qy        334 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 393
              |||| || |||  |    |  ||  ||  |  ||  |||||||||| || ||| || ||
Db        219 CTTGTTTTCTCCTTTCTTTCTTTCTTTTGTTTCTTTCCTGCTTGCTTTCTTGCTTGCTTG 278

Qy        394 CTTGCTTGCTCGCTGGCCTGCTTGCT 419
              ||| | ||||  || || | |||| |
Db        279 CTTTCGTGCTTTCTTGCTTTCTTGTT 304


RESULT 23
US-09-764-891-10047
; Sequence 10047, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC006
;  CURRENT APPLICATION NUMBER: US/09/764,891
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 10231
```

```
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10047
;   LENGTH: 368
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-891-10047

  Query Match             7.7%;  Score 77.2;  DB 11;  Length 368;
  Best Local Similarity   64.1%;  Pred. No. 4.2e-12;
  Matches  132;  Conservative    0;  Mismatches   73;  Indels    1;  Gaps    1;

Qy        215 AGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 274
              | |  ||||  |||||| |   |  ||||  |  || | ||||||| | || |  | |||
Db         99 ATGCTTGCTTTCTTGCTTGCTTGCTTGCTTTCGTGCTTTCTTGCTTTCCCGTTTTCTTGC 158

Qy        275 TTGCTTGCTCGC-TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 333
              || ||| ||  | |    | ||| |||||| |||  | ||||||||||||  |  || |
Db        159 TTTCTTTCTTTCTTTTGTTTCTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTT 218

Qy        334 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 393
              |||| || |||  |    |  ||  ||  |  ||  ||||||||||| || ||| || ||
Db        219 CTTGTTTTCTCCTTTCTTTCTTTCTTTTGTTTCTTTCCTGCTTGCTTTCTTGCTTGCTTG 278

Qy        394 CTTGCTTGCTCGCTGGCCTGCTTGCT 419
              ||| | ||||  || || | |||| |
Db        279 CTTTCGTGCTTTCTTGCTTTCTTGTT 304


RESULT 24
US-09-728-446-736/c
; Sequence 736, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 736
;   LENGTH: 392
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(392)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-736

  Query Match             7.7%;  Score 77.2;  DB 9;  Length 392;
  Best Local Similarity   59.4%;  Pred. No. 4.3e-12;
  Matches  130;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
              | ||||  | |  |     ||||   |   |||||   |  | ||  | | ||| | | |
Db        223 CCTGCTGTTGGTCCCTGCTGCTGCATTTGCTGCTTTTGCTATTGCTGGATGGCTGGATGG 164

Qy        286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
              ||||  |||| | | ||| | ||||  | ||||| | | |||||   ||| | ||| | |
Db        163 CTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGGTGG 104

Qy        346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 405
                |||  | | ||| | | |||||  | || | | ||| | ||||  | ||| | | | |
Db        103 ATGGCTGGATGGCTGGATGGCTGGTTTNCTGGATGGCTGGATGGCTGGGTTGGTGGATGG 44

Qy        406 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 444
              ||||   ||| | |  | |||  || ||   ||| | ||
Db         43 CTGGATGGCTGGGTTGGTGGATTGTTGGATAGCTGGTTT 5


RESULT 25
US-09-728-446-325/c
; Sequence 325, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 325
;   LENGTH: 410
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(410)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-325

  Query Match             7.6%;  Score 75.8;  DB 9;  Length 410;
  Best Local Similarity   59.3%;  Pred. No. 1.1e-11;
  Matches  128;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy        226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
              | ||||  | |  |     ||||   |   |||||   |  | ||  | | ||| | | |
Db        241 CCTGCTGTTGGTCCCTGCTGCTGCATTTGCTGCTTTTGCTATTGCTGGATGGCTGGATGG 182

Qy        286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
              ||||  |||| | | ||| | ||||  | ||||| | | |||||   ||| | ||| | |
Db        181 CTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGGTGG 122

Qy        346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 405
               ||||  | | ||| | |·|||||  |||| | | ||| | ||||  | ||  | | | |
Db        121 ATGGCTGGATGGCTGGATGGCTGGTTTGCTGGATGGCTGGATGGCTGGGTTCGTGGATGG 62

Qy        406 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTG 441
              ||||   ||| | |  | |||  || ||   ||| | ·
Db         61 CTGGATGGCTGGGTTGGTGGATTGTTGGATNGCTGG 26


RESULT 26
US-09-728-446-250/c
; Sequence 250, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 250
;   LENGTH: 427
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
```

```
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(427)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-250

  Query Match             7.4%;  Score 74.4;  DB 9;  Length 427;
  Best Local Similarity   58.3%;  Pred. No. 2.6e-11;
  Matches  126;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy        226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
              | ||||  | |  |     ||||   |    ||||  |  | ||  | | ||| | |
Db        258 CCTGCTGTTGGTCCCTGCTGCTGCATTTGCNGCTTTTGCTATTGCTGGATGGCTGGATGN 199

Qy        286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
              ||||  |||| | | ||| | ||||  | |||||  | | |||||   ||| | ||| |
Db        198 CTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGGTGN 139

Qy        346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 405
              ||||  | | ||| | | |||||  ||||   | ||| | ||||  | ||| | | | |
Db        138 ATGGCTGGATGGCTGGATGGCTGGTTTGCTGNATGGCTGGATGGCTGGGTTGGTGGATGG 79

Qy        406 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTG 441
              ||||   ||| | |  | |||  || ||  |||| |
Db         78 CTGGATGGCTGGGTTGGTGGATTGTTGGATTGCTGG 43


RESULT 27
US-10-027-632-273001/c
; Sequence 273001, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 273001
;   LENGTH: 1006
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-273001

  Query Match             7.4%;  Score 73.8;  DB 13;  Length 1006;
  Best Local Similarity   45.6%;  Pred. No. 4.4e-11;
  Matches  297;  Conservative    0;  Mismatches  352;  Indels    2;  Gaps    1;

Qy         97 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 156
              | |||  |||  |  ||   | | | | | |||   | | | ||   |||  |||||
Db        924 GATGAATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGTAGGATG 865

Qy        157 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAG 216
              | | | ||   |||  |  |    | | | | | |||   | | | ||   |||  |  |
Db        864 GATGGATGGATGGATGGATGATAGGATGGATGGATGGGTGGATGGATGGTAGGATGGATG 805

Qy        217 GGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 276
              |   | |||  | | | |||||   | | | |||   | | | | | | | |||   | |
Db        804 GTAGGATGGATGGATGGTAGGATGGATGGATGGATGGATGGATAGGTGGATAGGTG 745

Qy        277 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 336
              | | | | | |||   | | |   |||      |   | |   | | | | |||   | |
Db        744 GATAGATAGGTGGATGGATGGGCAGCTAAGGTGGTAGGTGAGTGGATGGATGGATGGATG 685

Qy        337 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 396
              | | | | | |||   | | | | | | | |     | | | | | | | |||   | |
Db        684 GATGGATGGATGGATGGATGGATGGATAGGTAAGTGGATGGRTGGATGGATGGATAGGTA 625

Qy        397 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACG 456
              | | | | |||   | | | |   | |        ||   | |   | | |    |   |
Db        624 AGTGGATGGGTGGATAGATGGATAGGTGTTTGAGTGGATGGATAAATAGATGTCTGAGTG 565

Qy        457 GTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTG--CTAGTAGGGCTGCTGGG 514
              |          || ||| | ||| |   ||| ||||||  |   ||  | ||     |||
Db        564 GATGAATGCATTGGTGGGTGGATGGGTAGGCAGCTGGGTAGATGTATGAAGGAGTGAGGG 505

Qy        515 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGG 574
              | |   ||     |    || || || |   ||  ||||   ||    |     |
Db        504 AGTTTGGATAGGTATACGAAAGGCCAGCAGAGAGCCTTCGCTGGATGTTTCTCAGTGAGT 445

Qy        575 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGG 634
                | | |     |||       | | || |    || ||   | | || ||   ||
Db        444 AGATGGATTGTGGGAGAGGGAGAGATGAATGGATGGATGGATGGATGGATGGATGGATGG 385

Qy        635 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGC 694
              ||   | | || || | || ||   | | || ||   || || ||| | || ||  ||
Db        384 ATGGATGGATGGATGGTAGGATGGATGGATGGATGGATGGATGATAGGATGGATGGATGG 325

Qy        695 TGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTG 745
                |  ||  || | |  ||   |  ||  |  ||  || ||  | || ||
Db        324 GTGGATGGATGGTAGGATGGATGGTAGGATGGATGGATGGTAGGATGGATG 274


RESULT 28
US-09-728-446-246/c
; Sequence 246, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 246
;   LENGTH: 379
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(379)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-246

  Query Match             7.3%;  Score 72.8;  DB 9;  Length 379;
  Best Local Similarity   57.9%;  Pred. No. 7.2e-11;
  Matches  125;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy        226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
              | ||||  | |  |     ||||   |   |||||   |  | ||  | | ||| | | |
Db        239 CCTGCTGTTGGTCCCTGCTGCTGCATTTGCTGCTTTTGCTATTGCTGGATGGCTGGATGG 180

Qy        286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
              ||||  |||| | | ||| | ||||  | |||||  | | |||||   ||| |  || |
Db        179 CTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGNTGGTGN 120
```

```
Qy        346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 405
              || |    | ||| | | |||||  |||| | | ||| | ||||  | ||| | | | |
Db        119 ATGNCTGNATGGCTGGATGGCTGGTTTGCTGGATGGCTGGATGGCTGGGTTGGTGGATGG 60

Qy        406 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTG 441
              ||||   ||| | |  | |||  ||  |  |||| |
Db         59 CTGGATGGCTGGGTTGGTGGATTGTTAGATTGCTGG 24


RESULT 29
US-10-040-739-1102/c
; Sequence 1102, Application US/10040739
; Publication No. US20020173635A1
;   GENERAL INFORMATION:
;        APPLICANT: Jacobs, Kenneth
;                   McCoy, John
;                   LaVallie, Edward
;                   Racie, Lisa
;                   Merberg, David
;                   Treacy, Maurice
;                   Spaulding, Vikki
;        TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
;        NUMBER OF SEQUENCES: 1519
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genetics Institute, Inc.
;             STREET: 87 CambridgePark Drive
;             CITY: Cambridge
;             STATE: Massachusetts
;             COUNTRY: U.S.A
;             ZIP: 02140
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy Disk
;             COMPUTER: IBM PC Compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/040,739
;             FILING DATE: 07-Jan-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/036,520
;             FILING DATE: 03-JUN-1998
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Brown, Scott A.
;             REGISTRATION NUMBER: 32,724
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (617) 498-8224
;             TELEFAX: (617) 876-5851
;   INFORMATION FOR SEQ ID NO: 1102:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 368 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1102:
US-10-040-739-1102

  Query Match             7.3%;  Score 72.6;  DB 13;  Length 368;
  Best Local Similarity   65.2%;  Pred. No. 8.1e-11;
  Matches  137;  Conservative    0;  Mismatches   67;  Indels    6;  Gaps    2;

Qy        220 TGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 279
              | || |||||||    .|  |||| ||| || | | ||||| || |||  |||| || ||
Db        227 TTCTTGCTTGCTTTCTTGCTTGCTNGCTTGCTTTCGTGCTTTCTTGCTTTCCTGTTTNCT 168

Qy        280 TGCTCGCTGGCCTGCTTG-----CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 334
              | ||  ||  | | |||       ||| || ||| || | |||||||||| |||  | |||
Db        167 TTCTTTCTTTCTTTCTTTTGTTTCTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTCGTGC 108

Qy        335 TTGCTTGCTCGCTGGCCTGCTTGCTTGC-TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 393
              || |||| |  || || | ||| ||| | |   |  | | | ||||||||  || || ||
Db        107 TTTCTTGTTTTCTCGCNTTCTTTCTTTCTTTTGTTTCTTTCCTGCTTGCTTTCTTGCTTG 48

Qy        394 CTTGCTTGCTCGCTGGCCTGCTTGCTGAGG 423
              ||||| |  |||  | ||||| || |||
Db         47 ATTGCTTTCGTGCTTTCTTGCTTTCTTAGG 18


RESULT 30
US-10-027-632-141458/c
; Sequence 141458, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 141458
;   LENGTH: 681
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-141458

  Query Match             7.0%;  Score 69.6;  DB 13;  Length 681;
  Best Local Similarity   58.3%;  Pred. No. 6.1e-10;
  Matches  120;  Conservative    1;  Mismatches   85;  Indels    0;  Gaps    0;

Qy        211 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 270
              ||  | |||||  || |||   || ||| ||    ||  | | ||| ||| ||  ||  |
Db        590 CACCATGGCTGGAGGATTGAGGGTGGGGATGGGATCTTTCTTTCTTTCTTTCTTTCTTTC 531

Qy        271 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 330
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||| ||  |
Db        530 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTCTCTC 471

Qy        331 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 390
              ||  ||: ||| ||  |   ||| ||| ||| ||  |   ||  ||| ||  ||  |
Db        470 TCTCTCTCTYTCTCTCTCTCTCTCTTTCTTTCTCTCTCTCTCTCTCTTCCTTCCTTCC 411

Qy        391 CTGCTTGCTTGCTCGCTGGCCTGCTT 416
              | ||| ||| ||  ||  | | |||
Db        410 TTCCTTCCTTTCTTCCTTCCTTCCTT 385


RESULT 31
US-10-027-632-141457/c
; Sequence 141457, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
```

```
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 141457
;   LENGTH: 681
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-141457

  Query Match             6.8%;  Score 68.4;  DB 13;  Length 681;
  Best Local Similarity   58.3%;  Pred. No. 1.3e-09;
  Matches  120;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        211 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 270
              ||  | |||||  || |||   || ||| ||    ||  | | ||| ||| ||  ||  |
Db        590 CACCATGGCTGGAGGATTGAGGGTGGGGATGGGATCTTTCTTTCTTTCTTTCTTTCTTTC 531

Qy        271 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 330
               | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||| ||  |
Db        530 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTCTCTC 471

Qy        331 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 390
                ||  ||  ||| ||  |   ||| ||| ||| ||  |   ||  ||| ||  ||  |
Db        470 TCTCTCTCTCTCTCTCTCTCTCTTTCTTTCTCTCTCTCTCTCTCTTCCTTCCTTCC 411

Qy        391 CTGCTTGCTTGCTCGCTGGCCTGCTT 416
               | ||| ||| ||  ||  | | |||
Db        410 TTCCTTCCTTTCTTCCTTCCTTCCTT 385


RESULT 32
US-10-027-632-141460/c
; Sequence 141460, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 141460
;   LENGTH: 681
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-141460

  Query Match             6.8%;  Score 68.4;  DB 13;  Length 681;
  Best Local Similarity   58.3%;  Pred. No. 1.3e-09;
  Matches  120;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        211 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 270
              ||  | |||||  || |||   || ||| ||    ||  | | ||| ||| ||  ||  |
Db        590 CACCATGGCTGGAGGATTGAGGGTGGGGATGGGATCTTTCTTTCTTTCTTTCTTTCTTTC 531

Qy        271 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 330
               | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||: ||| ||  |
Db        530 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTYTCTCTCTC 471

Qy        331 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 390
                ||  ||  ||| ||  |   ||| ||| ||| ||  |   ||  ||| ||  ||  |
Db        470 TCTCTCTCTCTCTCTCTCTCTCTTTCTTTCTCTCTCTCTCTCTCTTCCTTCCTTCC 411

Qy        391 CTGCTTGCTTGCTCGCTGGCCTGCTT 416
               | ||| ||| ||  ||  | | |||
Db        410 TTCCTTCCTTTCTTCCTTCCTTCCTT 385


RESULT 33
US-10-027-632-141459/c
; Sequence 141459, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 141459
;   LENGTH: 681
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-141459

  Query Match             6.8%;  Score 68;  DB 13;  Length 681;
  Best Local Similarity   57.8%;  Pred. No. 1.7e-09;
  Matches  119;  Conservative    1;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        211 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 270
              ||  | |||||  || |||   || ||| ||    ||  | | ||| ||| ||  ||  |
Db        590 CACCATGGCTGGAGGATTGAGGGTGGGGATGGGATCTTTCTTTCTTTCTTTCTTTCTTTC 531

Qy        271 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 330
               | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||: ||| ||  |
Db        530 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTYTCTCTCTC 471

Qy        331 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 390
                ||  ||  ||| ||  |   ||| ||| ||| ||  |   ||  ||| ||  ||  |
Db        470 TCTCTCTCTCTCTCTCTCTCTCTTTCTTTCTCTCTCTCTCTCTCTTCCTTCCTTCC 411

Qy        391 CTGCTTGCTTGCTCGCTGGCCTGCTT 416
               | ||| ||| ||  ||  | | |||
Db        410 TTCCTTCCTTTCTTCCTTCCTTCCTT 385


RESULT 34
```

```
US-09-728-446-465
; Sequence 465, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 465
;   LENGTH: 397
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(397)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-465

  Query Match             6.7%;  Score 67;  DB 9;  Length 397;
  Best Local Similarity   65.5%;  Pred. No. 3e-09;
  Matches   97;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy        242 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 301
              ||| |||   | | || |  | |  |||  |||| | ||||| | | ||  | ||||| |
Db         32 CTCCCTGTTGTTCCTGTTGCCACTTTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGG 91

Qy        302 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 361
               | |||||  |||| | ||||| | | ||  | ||||| | | |||||  |||| | |||
Db         92 ATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATNGCTGGGTTGCTGGGTTG 151

Qy        362 CTCGCTGGCCTGCTTGCTTGCTCGCTGG 389
              || | | ||  | ||||| | | |||||
Db        152 CTGGATTGCTGGATTGCTGGATTGCTGG 179


RESULT 35
US-09-764-877-3185
; Sequence 3185, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC005
;  CURRENT APPLICATION NUMBER: US/09/764,877
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 4031
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3185
;   LENGTH: 223
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-877-3185

  Query Match             6.6%;  Score 65.8;  DB 10;  Length 223;
  Best Local Similarity   60.2%;  Pred. No. 6e-09;
  Matches  109;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
              ||  | |||    ||  |||||  | |||    ||  |||||| | |||    |   |||
Db         43 CTCTTAGCTCAGTTGGGTGCTTAGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTAGGTGC 102

Qy        299 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 358
              ||| | |||    || ||||||  | |||    || ||| | ||||  |    ||||||
Db        103 TTGGTAGCTCAGTTGGGTGCTTGGTAGCTCAGGTGTTTGGTAGCTCAGTTAGGTGCTTGG 162

Qy        359 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 418
              | ||||  | |  ||||||  | ||||  | |  ||||||  | ||||  | |  ||||||
Db        163 TAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGA 222

Qy        419 T 419
              |
Db        223 T 223


RESULT 36
US-09-728-446-978/c
; Sequence 978, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 978
;   LENGTH: 393
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(393)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-978

  Query Match             6.5%;  Score 65.2;  DB 9;  Length 393;
  Best Local Similarity   54.5%;  Pred. No. 9.5e-09;
  Matches  115;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
              | ||||  | |  |  |  ||||   |   |||||  |    ||  | | ||| | | |
Db        224 CCTGCTGTTGGTCCCTCCTGCTGCATTTNATGCTTTTGCTANTGCTGGATGGCTGGATGG 165

Qy        286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
              ||||  ||||   | ||| | ||||  | |||||  | |  ||||   ||| | ||| |
Db        164 CTGGGTTGCTGNATGGCTGGATGGCTGGGTTGCTGGATGCCTGGATGGCTGGGTTGGTGN 105

Qy        346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 405
               ||||    |  || | | |||||  ||||   |  || | ||||  |  ||     | |
Db        104 ATGGCTGNATGNCTGGATGGCTGGTTTGCTGNATGNCTGGATGGCTGGGNTGGNCNATGG 45

Qy        406 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 436
              ||||    || | |  | |||  || |  ||
Db         44 CTGGATGNCTGGGTTGGTGGATTGTTGNACT 14


RESULT 37
US-10-017-161-1581
; Sequence 1581, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
```

```
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1581
;   LENGTH: 13830
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(13830)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(486)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (4460)..(4952)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (5748)..(5918)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (6811)..(6954)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (10095)..(10238)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (10541)..(10707)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (13494)..(13630)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (2697)..(2796)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (13826)..(13830)
;   OTHER INFORMATION: a, t, c, g, unknown or other
US-10-017-161-1581

  Query Match             6.5%;  Score 65.2;  DB 12;  Length 13830;
  Best Local Similarity   52.2%;  Pred. No. 1.6e-08;
  Matches  145;  Conservative    0;  Mismatches  133;  Indels    0;  Gaps    0;

Qy        136 TGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 195
              |||| |||   | |   ||||||   |        | ||   ||| | |||  ||   ||
Db       4639 TGCTCAGGTACCTGCCCGCCTGCCCACCTGCCCACCTGTCCACCTACCTGCCTGCCCACC 4698

Qy        196 TGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCT 255
              |||   | |        ||    ||||   | |||      | |  |  |    ||| |
Db       4699 TGCCCACCGGCCTACTTGTCCACCTGCATACCTGCCCACCTGCCCACCTGTCCACCTACC 4758

Qy        256 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 315
               || ||||| || ||||||  ||| | |  |    ||||| | | |||  ||  ||||||
Db       4759 CGCCTGCTCACTTGCCTGCCTGCCTACCTGTCCACCTGCATACCTGCCCGCCTACCCACC 4818

Qy        316 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 375
              || |  |   |  ||||||   | || || |   ||||| | | |||  ||  ||||||
Db       4819 TGTTCACCTACCTGCCTGCCCACCTGTTCACCTACCTGCCTCCCTGCCGGCCTGCCTGCC 4878

Qy        376 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 413
              | | ||   ||| | |||  ||  | |   |||  | |
Db       4879 TACCTGTCTGCTTGTCTGTCTGTATTCAGCCTGCTCAG 4916


RESULT 38
US-09-764-872-812
; Sequence 812, Application US/09764872
; Publication No. US20030050231A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA125
;  CURRENT APPLICATION NUMBER: US/09/764,872
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 957
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 812
;   LENGTH: 31718
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-872-812

  Query Match             6.5%;  Score 65.2;  DB 11;  Length 31718;
  Best Local Similarity   62.0%;  Pred. No. 1.8e-08;
  Matches  103;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        246 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 305
              ||  | | ||||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy        306 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy        366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              ||       |   |  ||| ||      ||  ||  ||| ||  ||
Db      15036 CTTCTTCTTTCTTTCTCTCTCTTTTTCTCTGTCTCTCTCTCTCTCC 15081


RESULT 39
US-09-764-872-813
; Sequence 813, Application US/09764872
; Publication No. US20030050231A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA125
;  CURRENT APPLICATION NUMBER: US/09/764,872
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 957
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 813
;   LENGTH: 31718
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-872-813

  Query Match             6.5%;  Score 65.2;  DB 11;  Length 31718;
  Best Local Similarity   62.0%;  Pred. No. 1.8e-08;
  Matches  103;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        246 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 305
              ||  | | ||||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy        306 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy        366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              ||       |   |  ||| ||      ||  ||  ||| ||  ||
Db      15036 CTTCTTCTTTCTTTCTCTCTCTTTTTCTCTGTCTCTCTCTCTCTCC 15081


RESULT 40
US-09-764-891-9103
; Sequence 9103, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC006
;  CURRENT APPLICATION NUMBER: US/09/764,891
```

```
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 10231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9103
;   LENGTH: 31718
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-891-9103

  Query Match             6.5%;  Score 65.2;  DB 11;  Length 31718;
  Best Local Similarity   62.0%;  Pred. No. 1.8e-08;
  Matches  103;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        246 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 305
              ||  | | |||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy        306 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy        366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              ||       |   | ||| ||      ||  ||  ||| ||  ||
Db      15036 CTTCTTCTTTCTTTCTCTCTCTTTTTCTCTGTCTCTCTCTCTCTCC 15081


RESULT 41
US-09-764-891-9104
; Sequence 9104, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC006
;  CURRENT APPLICATION NUMBER: US/09/764,891
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 10231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9104
;   LENGTH: 31718
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-891-9104

  Query Match             6.5%;  Score 65.2;  DB 11;  Length 31718;
  Best Local Similarity   62.0%;  Pred. No. 1.8e-08;
  Matches  103;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        246 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 305
              ||  | | |||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy        306 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy        366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              ||       |   | ||| ||      ||  ||  ||| ||  ||
Db      15036 CTTCTTCTTTCTTTCTCTCTCTTTTTCTCTGTCTCTCTCTCTCTCC 15081


RESULT 42
US-10-027-632-281951/c
; Sequence 281951, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 281951
;   LENGTH: 647
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-281951

  Query Match             6.5%;  Score 65;  DB 13;  Length 647;
  Best Local Similarity   64.6%;  Pred. No. 1.2e-08;
  Matches  113;  Conservative    0;  Mismatches   60;  Indels    2;  Gaps    1;

Qy        245 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 304
              ||||  |||||||||||||| ||| |||||| ||| |||  ||  || |||| ||| |||
Db        409 GCTGAGCTGCTTGCTTGCTTGCCTGCCTGCCTGCCTGCCTGCTTGCTTTCTTTCTC 350

Qy        305 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 364
              ||  | |  || ||| ||  ||  | | ||| ||| ||  ||  | | ||| | | ||
Db        349 TCTTTCTT--TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCCTTCTT 292

Qy        365 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              ||  | | ||| ||  ||  ||  | | ||| ||| ||| ||  | | ||  ||
Db        291 TCTTTCTTTCTTTCTCTCTTTCTTTCTTTCTTTCTTTCTCTCTTTCTTTCTCTCT 237


RESULT 43
US-10-027-632-273402
; Sequence 273402, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 273402
;   LENGTH: 495
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-273402

  Query Match             6.5%;  Score 64.6;  DB 13;  Length 495;
  Best Local Similarity   58.0%;  Pred. No. 1.4e-08;
```

```
  Matches  112;  Conservative    1;  Mismatches   80;  Indels    0;  Gaps    0;

Qy        225 GCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 284
              | ||| | | | ||  | | | |||  || ||| ||| | | | |  || ||| ||| |
Db        282 GATTGTTTGGATGGTAGGTTGGTGGATTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTG 341

Qy        285 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 344
              | | |  || | | ||| | | | |  || | | ||| | | | |  || ||| ||| |
Db        342 GGTTGGTTGGTGGGTTGGTTGGTTGGTTGGTGGGTTGGTTGGTTGGTTGGTTGGTT 401

Qy        345 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 404
              | ||   || ||| |||:| | | |  || ||| ||| | | | |   | ||  ||| |
Db        402 GGTGAGTTGGTTGGTTGSTTGGTTGGTTGGTTGGTTGGTGGGTTGGTGGGTTTGTTGGTT 461

Qy        405 GCTGGCCTGCTTG 417
              | ||   || |||
Db        462 GGTGAGTTGGTTG 474


RESULT 44
US-10-175-523-80
; Sequence 80, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
;  APPLICANT: Brockman, Jeffrey
;  APPLICANT:  Evans, David
;  APPLICANT:  Hook, Derek
;  APPLICANT:  Klimczak, Leszek
;  APPLICANT:  Laeng, Pascal
;  APPLICANT:  Palfreyman, Michael
;  APPLICANT:  Rajan, Prithi
;  TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
;  FILE REFERENCE: 3235/1J795-US3
;  CURRENT APPLICATION NUMBER: US/10/175,523
;  CURRENT FILING DATE:  2002-06-18
;  PRIOR APPLICATION NUMBER: US 60/299,151
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: US 60/317,828
;  PRIOR FILING DATE: 2001-09-07
;  PRIOR APPLICATION NUMBER: US 60/325,150
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 60/333,047
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: US 60/349,936
;  PRIOR FILING DATE: 2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/361,834
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 197
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 80
;   LENGTH: 90650
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-175-523-80

  Query Match             6.4%;  Score 64.4;  DB 14;  Length 90650;
  Best Local Similarity   56.7%;  Pred. No. 3.5e-08;
  Matches  119;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy        240 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 299
              | ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||
Db       2647 TCCTTTCTTCCTTTCTTCCTTTCTTCCTTTCTTCCTTTCTTCCTTTCTTCCTTTCTTCCT 2706

Qy        300 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 359
              | ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||
Db       2707 TTCTTCCTTCCTTCCTTTCTTCCTTCCTTTCTTCCTTTCTTCCTTTCTTCCTTTCTTCCT 2766

Qy        360 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              | |  ||  | | ||| ||| ||  |  | | ||| ||| ||  ||  | | ||| ||
Db       2767 TCTTTCCTTTCTTCCTTTCTTCCTTCTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 2826

Qy        420 GAGGGGACAGTAGGGCTGCTTGCTTGCTAA 449
                |  || | ||    |  || | |  ||

Db       2827 TATTAGATATTATCTTTATTTACATTTCAA 2856


RESULT 45
US-09-764-877-3186
; Sequence 3186, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC005
;  CURRENT APPLICATION NUMBER: US/09/764,877
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 4031
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3186
;   LENGTH: 223
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-877-3186

  Query Match             6.4%;  Score 64.2;  DB 10;  Length 223;
  Best Local Similarity   59.7%;  Pred. No. 1.7e-08;
  Matches  108;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
              ||  | |||    ||  |||||  | |||    ||  |||||| | |||    |  |||
Db         43 CTCTTAGCTCAGTTGGGTGCTTAGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTAGGTGC 102

Qy        299 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 358
              ||| | |||    ||  |||||| | |||    || ||| | ||||  |    || |||
Db        103 TTGGTAGCTCAGTTGGGTGCTTGGTAGCTCAGGTGTTTGGTAGCTCAGTTAGGTGTTTGG 162

Qy        359 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 418
              | ||||  | |  |||||| | ||||  | |  |||||| | ||||  | |  ||||||
Db        163 TAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGA 222

Qy        419 T 419
              |
Db        223 T 223


RESULT 46
US-09-089-818B-6/c
; Sequence 6, Application US/09089818B
; Patent No. US20020164682A1
;  GENERAL INFORMATION:
;    APPLICANT:  Follettie, Maximillian
;    APPLICANT:  DeRobertis, Edward M.
;    TITLE OF INVENTION:  Mammalian Cerberus-Like Protein &
;    TITLE OF INVENTION:  Compositions
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/089,818B
;      FILING DATE:  June 3, 1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LAZAR, STEVEN R
;      REGISTRATION NUMBER:  32,618
;      REFERENCE/DOCKET NUMBER:  GI 5290A
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8260
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3595 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  117..623
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  624..2402
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  2403..2699
US-09-089-818B-6

  Query Match             6.3%;  Score 63.4;  DB 10;  Length 3595;
  Best Local Similarity   61.9%;  Pred. No. 4.2e-08;
  Matches  117;  Conservative    0;  Mismatches   71;  Indels    1;  Gaps    1;

Qy        222 CTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 281
              ||  | | | | ||   |   | ||  |||| | ||| |||  ||  ||||||| ||| ||
Db       3303 CTTTCCTACAAATATTTCACGTAGCCTGCCTCCCTGCCTGCCTGCGTGCCTGCCTGCCTG 3244

Qy        282 CTCGCTGGCCTGCTTG-CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 340
              |  ||  |||||| || ||| ||  ||  | | ||| ||| ||  ||  | | ||| |||
Db       3243 CCTGCCTGCCTGCCTGCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTT 3184

Qy        341 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
              ||  ||  | |  || ||| ||  ||  | | ||| ||  ||| ||  | ||||  ||
Db       3183 CCTTCCTTCCTTCTTTCCTTCCTTCCTTCCTTCCTTCCTCCCTCTCTATCTTGCTTCTTT 3124

Qy        401 GCTCGCTGG 409
              ||| |||||
Db       3123 GCTTGCTGG 3115


RESULT 47
US-10-027-632-161036
; Sequence 161036, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 161036
;   LENGTH: 863
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-161036

  Query Match             6.3%;  Score 63.2;  DB 13;  Length 863;
  Best Local Similarity   63.2%;  Pred. No. 3.8e-08;
  Matches  129;  Conservative    0;  Mismatches   73;  Indels    2;  Gaps    2;

Qy        202 CTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTG 261
              |||| ||    |  ||   |||||||||||||| | |||||||||| | ||||| | |
Db        642 CTGATGGCTTGGCTGGCTGGCTGGCTTGCTAGCATGGCTGCTCGCTTGGCTGCTGGGTGC 701

Qy        262 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 321
              | | ||||  ||| |||   |  |||||  ||| |||   | |||||| ||||  || |
Db        702 TTGGGTGGCTGGCTGGCTGCTTTCCTGGCTGGCTGGCTGCATGGCTGGCTTGCTGACTAG 761

Qy        322 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 381
              | ||| | | |||||||||| | |||| |||   |||||| | ||         |
Db        762 GT-GCTTGACGGCTTGCTTGCT-GGTGGCTTGCCGCCTTGCTGGTTGTTGCTTGCCCGCC 819

Qy        382 CTCGCTGGCCTGCTTGCTTGCTCG 405
              |       |||||||| |||| |
Db        820 TTGTGGCTGCTGCTTGCCTGCTGG 843


RESULT 48
US-10-027-632-161037
; Sequence 161037, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 161037
;   LENGTH: 863
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-161037

  Query Match             6.3%;  Score 63.2;  DB 13;  Length 863;
  Best Local Similarity   63.2%;  Pred. No. 3.8e-08;
  Matches  129;  Conservative    0;  Mismatches   73;  Indels    2;  Gaps    2;

Qy        202 CTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTG 261
              |||| ||    |  ||   |||||||||||||| | |||||||||| | ||||| | |
Db        642 CTGATGGCTTGGCTGGCTGGCTGGCTTGCTAGCATGGCTGCTCGCTTGGCTGCTGGGTGC 701

Qy        262 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 321
              | | ||||  ||| |||   |  |||||  ||| |||   | |||||| ||||  || |
Db        702 TTGGGTGGCTGGCTGGCTGCTTTCCTGGCTGGCTGGCTGCATGGCTGGCTTGCTGACTAG 761

Qy        322 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 381
              | ||| | | |||||||||| | |||| |||   |||||| | ||         |
Db        762 GT-GCTTGACGGCTTGCTTGCT-GGTGGCTTGCCGCCTTGCTGGTTGTTGCTTGCCCGCC 819

Qy        382 CTCGCTGGCCTGCTTGCTTGCTCG 405
              |       |||||||| |||| |
Db        820 TTGTGGCTGCTGCTTGCCTGCTGG 843
```

```
RESULT 49
US-10-027-632-161038
; Sequence 161038, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 161038
;   LENGTH: 863
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-161038

  Query Match             6.3%;  Score 63.2;  DB 13;  Length 863;
  Best Local Similarity   63.2%;  Pred. No. 3.8e-08;
  Matches  129;  Conservative    0;  Mismatches   73;  Indels    2;  Gaps    2;

Qy        202 CTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTG 261
              |||| ||    |  ||   |||||||||||||| | ||||||||||| | ||||| | |
Db        642 CTGATGGCTTGGCTGGCTGGCTGGCTTGCTAGCATGGCTGCTCGCTTGGCTGCTGGGTGC 701

Qy        262 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 321
              | | ||||  ||| |||   |  |||||  ||| |||   | ||||| |||| || | |
Db        702 TTGGGTGGCTGGCTGGCTGCTTTCCTGGCTGGCTGGCTGCATGGCTGGCTTGCTGACTAG 761

Qy        322 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 381
              | ||| | | |||||||||||| | |||| |||   |||||| | ||         |
Db        762 GT-GCTTGACGGCTTGCTTGCT-GGTGGCTTGCCGCCTTGCTGGTTGTTGCTTGCCCGCC 819

Qy        382 CTCGCTGGCCTGCTTGCTTGCTCG 405
              |       |||||||| |||| |
Db        820 TTGTGGCTGCTGCTTGCCTGCTGG 843


RESULT 50
US-10-311-455-300
; Sequence 300, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 300
;   LENGTH: 6109
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 1763, 3313..3314, 5896
;   OTHER INFORMATION: n is a or g or c or t
US-10-311-455-300

  Query Match             6.3%;  Score 62.8;  DB 12;  Length 6109;
  Best Local Similarity   45.3%;  Pred. No. 6.6e-08;
  Matches  268;  Conservative    0;  Mismatches  322;  Indels    2;  Gaps    1;

Qy        105 GACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTG 164
              ||  | | |  ||  ||   |   |  ||  ||   ||  ||  |  ||   | | | ||
Db       3926 GAATGAATGGGTGGGTGGGTGAGTGGATGGGTGGGTAGATGATGGGTGGATGGGTGGGTG 3985

Qy        165 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTG 224
                | |||     |   | | | | | |||   | | | || | |   |  ||   | ||
Db       3986 GGTGGATGAGTGAATGGGTGGATGGATGGGTGGATAGATGGGTGAGTGGGTGGATAGATG 4045

Qy        225 GCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 284
              | |   |   | |  | | | |||   | | | | | | ||   | | | | | |
Db       4046 GGTGAGTGAGTGAGTGGGTGGATGGGTGGGTGGATGGATAGATGTGTGGATGGGTGGGTG 4105

Qy        285 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 344
              | |||   | | | |   | | ||   | | | |   | | |||   | |   | | |
Db       4106 GATGGATAGATGGGTGAATGGGCGGAAAGATGGGTGAATGGGTGGATGGATGAATGGATA 4165

Qy        345 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 404
              |||     | | | | | | |||   | | | |   | | |||   | | | | | |
Db       4166 AATGGGTAAATGGGTGGGTGGATGGAGAGATGGGTGAATGGGTGGGTAGGTGGATGGATG 4225

Qy        405 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCT 464
              | ||   | |   |||  ||   | | || ||  ||  ||    || |     | | |
Db       4226 GGTGTGTGGGTGAATGAATGG--TGGATGGGTGTGTGGGTGGGTGGGTGGATAGATGGGT 4283

Qy        465 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 524
              | ||   | ||   | | |  | |||  | |||  |||  | |||  || | | |  |
Db       4284 GAATGTGTGGATGGGTGGATAGAGGATGGGAGGATAGATGGGTGGATGGGTTGATGGATG 4343

Qy        525 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 584
              |   | |||   | |     |   | |||   | |    |||    | |   | |||  |
Db       4344 GATGGATGGATGGGTGTATAGATGGATGGATGGGTGAATGGGTGGGTAGATGGATAGATG 4403

Qy        585 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 644
              ||  | ||||  | |   ||     ||||  | | |  |||  | ||||  | | |  |
Db       4404 GGTGGATGGGTCGATGTATGGATGGATGGGTAGATGGGTGGGTGGATGGGTGGATGGGTG 4463

Qy        645 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTG 696
              |   | ||||  | | |  ||   | ||||  | | |   ||  |  |  ||
Db       4464 GATGGGTGGGTGGATGGGTGGATGGGTGGGTGGGTGGATAGGTGGATTAATG 4515


RESULT 51
US-10-175-523-96
; Sequence 96, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
;  APPLICANT: Brockman, Jeffrey
;  APPLICANT:  Evans, David
;  APPLICANT:  Hook, Derek
;  APPLICANT:  Klimczak, Leszek
;  APPLICANT:  Laeng, Pascal
;  APPLICANT:  Palfreyman, Michael
;  APPLICANT:  Rajan, Prithi
;  TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
;  FILE REFERENCE: 3235/1J795-US3
```

```
;  CURRENT APPLICATION NUMBER: US/10/175,523
;  CURRENT FILING DATE:  2002-06-18
;  PRIOR APPLICATION NUMBER: US 60/299,151
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: US 60/317,828
;  PRIOR FILING DATE: 2001-09-07
;  PRIOR APPLICATION NUMBER: US 60/325,150
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 60/333,047
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: US 60/349,936
;  PRIOR FILING DATE: 2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/361,834
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 197
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 96
;   LENGTH: 110079
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(110079)
;   OTHER INFORMATION: where n may be a or g or c or t/u, unknown, or other
US-10-175-523-96

  Query Match             6.3%;  Score 62.8;  DB 14;  Length 110079;
  Best Local Similarity   56.2%;  Pred. No. 1e-07;
  Matches  118;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy        228 TGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 287
              || || |     | ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||
Db     100474 TGATATTTCTTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 100533

Qy        288 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 347
              | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||
Db     100534 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTCCTTCCTTCCTTCCTTCCTTCCT 100593

Qy        348 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 407
              | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| |||
Db     100594 TCCTTCCTTCCTTCCTTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCAGA 100653

Qy        408 GGCCTGCTTGCTGAGGGGACAGTAGGGCTG 437
              |  | |||    |   |   || ||||
Db     100654 AATCCACCTGCCTCTGCCTCCTAAGTGCTG 100683


RESULT 52
US-10-027-632-174763/c
; Sequence 174763, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 174763
;   LENGTH: 2940917
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(2940917)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-174763

  Query Match             6.2%;  Score 62.2;  DB 13;  Length 2940917;
  Best Local Similarity   76.8%;  Pred. No. 2.4e-07;
  Matches   76;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
              ||||||||||||| ||| || |||||||||||| ||| || |||||||||||| ||| |||
Db    2504045 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCC 2503986

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 350
                | | | | ||| | |   ||| ||| |||||||  ||
Db    2503985 ACTTCGTTCGTTCGTTCGTTCGCTCGCTCGCTCGCTCGC 2503947


RESULT 53
US-10-027-632-15974
; Sequence 15974, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15974
;   LENGTH: 726
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-15974

  Query Match             6.2%;  Score 62;  DB 13;  Length 726;
  Best Local Similarity   48.1%;  Pred. No. 8.1e-08;
  Matches  205;  Conservative    1;  Mismatches  216;  Indels    4;  Gaps    1;

Qy        271 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 330
              ||||  |  ||  |      ||||| |  |||  || |  ||||  |   |  | | |
Db         22 CTGCAGGTATGGACATGCCTCTGCTGGTATGCATGCAGCTCTGCAGGTGGGGACACGGCT 81

Qy        331 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 390
              ||||  |   |  |   |  ||||  |  |||  ||  |  ||||  |   |  | | |
Db         82 CTGCAGGTGGGGACATGGCTCTGCAGGTATGCACGTGGCTCTGCAGGTGGGGACACGGCT 141

Qy        391 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAG 450
              ||||  |  |   |       |    |||||    |   | ||  ||||  |   :   |
Db        142 CTGCAGGTGTCGACATGCCTC----TGCTGGTATGCACGCAGCTCTGCAGGTGGRGACAT 197

Qy        451 GGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 510
              ||  | | | |  ||||||  |   ||   ||| |  ||  |  ||||  ||| |  |||
```

```
Db       198 GGCTCTGCAGGTATGCCTGCGGCTCTGCAGGTATGCATGTGGCTCTGCAGGTATGCATGC 257

Qy       511 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCC 570

Db       258 AGCTCTGCAGGTGGGGACATGCCTCTACCAGTATGCACGTGGCTCTGCAGGTGTGCACCC 317

Qy       571 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 630

Db       318 GGCTCTGCAGGTATGGACACAGCTCTGCAGGTGTGGACAGGGCTCTGCAGGTGTGCATGC 377

Qy       631 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 690

Db       378 AGCTCTACAGGTGTGGATGCACCTCTGCAGGTATGCACATGGCCCCTGGAGCTGTCTTCC 437

Qy       691 CTGCTG 696

Db       438 CTGCTG 443


RESULT 54
US-10-017-161-1673
; Sequence 1673, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1673
;   LENGTH: 13240
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(13240)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(261)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1594)..(2145)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2900)..(3044)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (3937)..(4080)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (7221)..(7364)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (7667)..(7833)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (10614)..(10746)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (12968)..(13040)
US-10-017-161-1673

  Query Match             6.2%;  Score 62;  DB 12;  Length 13240;
  Best Local Similarity   53.3%;  Pred. No. 1.2e-07;
  Matches  131;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy       200 TGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCT 259

Db      1773 TGCTCAGGTACCTGCCCGCCTGCCCACCTGCCCACCTGTCCACCTACCTGCCTGCCCACC 1832

Qy       260 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 319

Db      1833 TGCCCACCGGCCTACTTGTCCACCTGCATACCTGCCCACCTGCCCACCTACCCGCCTGCT 1892

Qy       320 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 379

Db      1893 CACTTGCCTGCCTGCCTACCTGTCCACCTGCATACCTGCCCGCCTACCCACCTGTTCACC 1952

Qy       380 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 439

Db      1953 TACCTGCCTGCCCACCTGTTCACCTACCTGCCTCCCTGCCGGCCTGCCTGCCTACCTGTC 2012

Qy       440 TGCTTG 445

Db      2013 TGCTTG 2018


RESULT 55
US-09-874-923-7/c
; Sequence 7, Application US/09874923
; Patent No. US20020081320A1
; GENERAL INFORMATION:
;  APPLICANT: Reed, Steven G.
;  APPLICANT:  Campos-Neto, Antonio
;  APPLICANT:  Webb, John R.
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Bhatia, Ajay
;  APPLICANT:  Coler, Rhea
;  APPLICANT:  Probst, Peter
;  APPLICANT:  Brannon, Mark
;  TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
;  TITLE OF INVENTION:  THERAPY AND DIAGNOSIS OF LEISHMANIASIS
;  FILE REFERENCE: 210121.420C8
;  CURRENT APPLICATION NUMBER: US/09/874,923
;  CURRENT FILING DATE:  2001-06-04
;  NUMBER OF SEQ ID NOS: 122
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 1771
;   TYPE: DNA
;   ORGANISM: Leishmania tropica
US-09-874-923-7

  Query Match             6.2%;  Score 61.8;  DB 9;  Length 1771;
  Best Local Similarity   48.1%;  Pred. No. 1e-07;
  Matches  295;  Conservative    0;  Mismatches  312;  Indels    6;  Gaps    4;

Qy       137 GCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 196

Db       961 GCTCCGCGCAGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGT 902

Qy       197 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 256

Db       901 GCATGGCCTCCAGGCGGGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGT 842

Qy       257 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 316

Db       841 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGC-T 783

Qy       317 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 376

Db       782 GCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCC 723

Qy       377 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC---CTGCTTGCTGAGGGGACAGTAGG 433

Db       722 GCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGACG 663

Qy       434 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 493
```

```
Db        662 CGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACA 603

Qy        494 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 553

Db        602 CGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGC 543

Qy        554 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 613

Db        542 TCCGCGCGGA-GACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTG 484

Qy        614 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 673

Db        483 CATGGCCTCCAGGCGGGCCGCCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-G 425

Qy        674 GCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTG 733

Db        424 CCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCG 365

Qy        734 CTACTAGGGCTGC 746

Db        364 CCTCCTCGGCTTC 352


RESULT 56
US-09-991-496-7/c
; Sequence 7, Application US/09991496
; Patent No. US20020169285A1
; GENERAL INFORMATION:
;  APPLICANT: Reed, Steven G.
;  APPLICANT:  Campos-Neto, Antonio
;  APPLICANT:  Webb, John R.
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Bhatia, Ajay
;  APPLICANT:  Coler, Rhea
;  APPLICANT:  Probst, Peter
;  APPLICANT:  Brannon, Mark
;  TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
;  TITLE OF INVENTION:  THERAPY AND DIAGNOSIS OF LEISHMANIASIS
;  FILE REFERENCE: 210121.420C9
;  CURRENT APPLICATION NUMBER: US/09/991,496
;  CURRENT FILING DATE:  2001-11-20
;  NUMBER OF SEQ ID NOS: 137
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 1771
;   TYPE: DNA
;   ORGANISM: Leishmania tropica
US-09-991-496-7

  Query Match             6.2%;  Score 61.8;  DB 10;  Length 1771;
  Best Local Similarity   48.1%;  Pred. No. 1e-07;
  Matches  295;  Conservative    0;  Mismatches  312;  Indels    6;  Gaps    4;

Qy        137 GCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 196

Db        961 GCTCCGCGCAGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGT 902

Qy        197 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 256

Db        901 GCATGGCCTCCAGGCGGGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGT 842

Qy        257 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 316

Db        841 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGC-T 783

Qy        317 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 376

Db        782 GCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCC 723

Qy        377 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC---CTGCTTGCTGAGGGGACAGTAGG 433

Db        722 GCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGACG 663

Qy        434 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 493

Db        662 CGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACA 603

Qy        494 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 553

Db        602 CGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGC 543

Qy        554 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 613

Db        542 TCCGCGCGGA-GACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTG 484

Qy        614 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 673

Db        483 CATGGCCTCCAGGCGGGCCGCCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-G 425

Qy        674 GCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTG 733

Db        424 CCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCG 365

Qy        734 CTACTAGGGCTGC 746

Db        364 CCTCCTCGGCTTC 352


RESULT 57
US-10-027-632-143381
; Sequence 143381, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 143381
;   LENGTH: 588
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-143381

  Query Match             6.1%;  Score 60.6;  DB 13;  Length 588;
  Best Local Similarity   48.4%;  Pred. No. 1.9e-07;
  Matches  168;  Conservative    0;  Mismatches  179;  Indels    0;  Gaps    0;

Qy        104 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 163

Db        162 GGATGGGTGGGTGGATCCATGGGTGGGTGTGTGGGTGGATGGACAGATGGGTTGATTGGT 221

Qy        164 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 223

Db        222 GGGTGAATGGATGGATGGTTGGATCCATGGGTAGGTGAGTGGATGGATTGGTGGATAGAT 281

Qy        224 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 283
```

```
Db        282 GGATGGATGGATGGATGGATGGATGGATGGATRGATGGATGGGTGAGTGGATAGATGGAT 341

Qy        284 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 343

Db        342 GGATGGATGAGCAGATGGTTGGGTTGATGGGTGGGTGAATGGATGGATGGTTGGATCCAT 401

Qy        344 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403

Db        402 GGGTAGGTGAGTTGATGGATTGCTGAGTGGGTGGGTGGATGGATGGATGGACGGATGGAT 461

Qy        404 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAG 450

Db        462 GGATAGATGGATTGGTGAGTGGATGGTTGGGTTGATGGATGGTTATG 508


RESULT 58
US-09-922-217-1058/c
; Sequence 1058, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Jiang, Yuqiu
;  APPLICANT:  Smith, Carole Lynn
;  APPLICANT:  King, Gordon E.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
;  TITLE OF INVENTION:  OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C13
;  CURRENT APPLICATION NUMBER: US/09/922,217
;  CURRENT FILING DATE:  2001-08-03
;  NUMBER OF SEQ ID NOS: 1124
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1058
;   LENGTH: 15720
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-922-217-1058

  Query Match             6.1%;  Score 60.6;  DB 9;  Length 15720;
  Best Local Similarity   45.8%;  Pred. No. 3.1e-07;
  Matches  280;  Conservative    0;  Mismatches  329;  Indels    2;  Gaps    2;

Qy        121 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCT 180

Db       5097 GCTGGGAGTGGTGGTTATTGGAGAGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGT 5038

Qy        181 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCT 240

Db       5037 GGTTATTGGAGGGCTGGGAGTGGTGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGG 4978

Qy        241 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 300

Db       4977 GGTGGTTGTGGTGGTTGGTGGAGGGCTGGGAGTGGTGGTTGGTGGAAGGGTGGTAGTGCT 4918

Qy        301 GCTCGCTGGCCTGCTTGCT-TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 359

Db       4917 GGTTGGTGGAGTGATGGGTGTAGTCGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGT 4858

Qy        360 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419

Db       4857 GGTTGTGGTCGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGAT 4798

Qy        420 GAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGAT 479

Db       4797 TGTTGGAGGACTGGGAGTGGTGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGT 4738

Qy        480 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 539

Db       4737 GGTTGTGGTGGTTGGTGGAGGGCTGGGAGTGGTGGTTGGTGGAAGGGTAGTAGTGCTGGT 4678

Qy        540 AGTAGGGCTGCTGGGC-TGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 598

Db       4677 TGGTGGAGTGATGGGCGTAGTCGTTGGAGGACTGGGAGTGGTGGTTGGTGGAGGGGTGGT 4618

Qy        599 TAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 658

Db       4617 TGTGGTGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAAGGGTGGTAGTGCTGGCTGG 4558

Qy        659 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGT 718

Db       4557 TGGAGTGATGGGCGTAGTCATTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGT 4498

Qy        719 GGGGCCGCTTG 729

Db       4497 GGTGGTTGTTG 4487


RESULT 59
US-09-833-263-1058/c
; Sequence 1058, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Meagher, Madeleine J.
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C12
;  CURRENT APPLICATION NUMBER: US/09/833,263
;  CURRENT FILING DATE:  2001-04-10
;  NUMBER OF SEQ ID NOS: 1093
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1058
;   LENGTH: 15720
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-833-263-1058

  Query Match             6.1%;  Score 60.6;  DB 10;  Length 15720;
  Best Local Similarity   45.8%;  Pred. No. 3.1e-07;
  Matches  280;  Conservative    0;  Mismatches  329;  Indels    2;  Gaps    2;

Qy        121 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCT 180

Db       5097 GCTGGGAGTGGTGGTTATTGGAGAGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGT 5038

Qy        181 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCT 240

Db       5037 GGTTATTGGAGGGCTGGGAGTGGTGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGG 4978

Qy        241 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 300

Db       4977 GGTGGTTGTGGTGGTTGGTGGAGGGCTGGGAGTGGTGGTTGGTGGAAGGGTGGTAGTGCT 4918

Qy        301 GCTCGCTGGCCTGCTTGCT-TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 359

Db       4917 GGTTGGTGGAGTGATGGGTGTAGTCGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGT 4858

Qy        360 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419

Db       4857 GGTTGTGGTCGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGAT 4798

Qy        420 GAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGAT 479

Db       4797 TGTTGGAGGACTGGGAGTGGTGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGT 4738

Qy        480 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 539
```

```
Db       4737 GGTTGTGGTGGTTGGTGGAGGGCTGGGAGTGGTGGTTGGTGGAAGGGTAGTAGTGCTGGT 4678

Qy        540 AGTAGGGCTGCTGGGC-TGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 598

Db       4677 TGGTGGAGTGATGGGCGTAGTCGTTGGAGGACTGGGAGTGGTGGTTGGTGGAGGGGTGGT 4618

Qy        599 TAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 658

Db       4617 TGTGGTGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAAGGGTGGTAGTGCTGGCTGG 4558

Qy        659 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGT 718

Db       4557 TGGAGTGATGGGCGTAGTCATTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGT 4498

Qy        719 GGGGCCGCTTG 729

Db       4497 GGTGGTTGTTG 4487


RESULT 60
US-10-025-380-1058/c
; Sequence 1058, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Jiang, Yuqiu
;  APPLICANT:  Smith, Carole L.
;  APPLICANT:  King, Gordon E.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  APPLICANT:  Skeiky, Yasir A. W.
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick Thomas S.
;  APPLICANT:  Carter, Darrick
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
;  TITLE OF INVENTION:  OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C14
;  CURRENT APPLICATION NUMBER: US/10/025,380
;  CURRENT FILING DATE:  2001-12-19
;  NUMBER OF SEQ ID NOS: 1129
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1058
;   LENGTH: 15720
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-025-380-1058

  Query Match             6.1%;  Score 60.6;  DB 13;  Length 15720;
  Best Local Similarity   45.8%;  Pred. No. 3.1e-07;
  Matches  280;  Conservative    0;  Mismatches  329;  Indels    2;  Gaps    2;

Qy        121 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCT 180

Db       5097 GCTGGGAGTGGTGGTTATTGGAGAGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGT 5038

Qy        181 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCT 240

Db       5037 GGTTATTGGAGGGCTGGGAGTGGTGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGG 4978

Qy        241 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 300

Db       4977 GGTGGTTGTGGTGGTTGGTGGAGGGCTGGGAGTGGTGGTTGGTGGAAGGGTGGTAGTGCT 4918

Qy        301 GCTCGCTGGCCTGCTTGCT-TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 359

Db       4917 GGTTGGTGGAGTGATGGGTGTAGTCGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGT 4858

Qy        360 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419

Db       4857 GGTTGTGGTCGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGAT 4798

Qy        420 GAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGAT 479

Db       4797 TGTTGGAGGACTGGGAGTGGTGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGT 4738

Qy        480 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 539

Db       4737 GGTTGTGGTGGTTGGTGGAGGGCTGGGAGTGGTGGTTGGTGGAAGGGTAGTAGTGCTGGT 4678

Qy        540 AGTAGGGCTGCTGGGC-TGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 598

Db       4677 TGGTGGAGTGATGGGCGTAGTCGTTGGAGGACTGGGAGTGGTGGTTGGTGGAGGGGTGGT 4618

Qy        599 TAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 658

Db       4617 TGTGGTGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAAGGGTGGTAGTGCTGGCTGG 4558

Qy        659 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGT 718

Db       4557 TGGAGTGATGGGCGTAGTCATTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGT 4498

Qy        719 GGGGCCGCTTG 729

Db       4497 GGTGGTTGTTG 4487


RESULT 61
US-10-311-455-762/c
; Sequence 762, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;. APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Det
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 762
;   LENGTH: 6171
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-762

  Query Match             6.0%;  Score 60.4;  DB 12;  Length 6171;
  Best Local Similarity   60.2%;  Pred. No. 3.1e-07;
  Matches  100;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy        242 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 301

Db       5814 CTCTCTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 5755

Qy        302 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 361

Db       5754 CTTTCTTTCCTTCTTTCTTTCTTTCTTCCTTTCTTTCTTTCTCTCTCTCTTTCTTCCTTC 5695

Qy        362 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 407

Db       5694 CTTCCTTCCCTCCCTCCCTCCCTCCCTCCTTCCTTCCTTCCTTCCT 5649
```

```
RESULT 62
US-10-027-632-143382
; Sequence 143382, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 143382
;   LENGTH: 588
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-143382

  Query Match             6.0%;  Score 60.2;  DB 13;  Length 588;
  Best Local Similarity   48.1%;  Pred. No. 2.5e-07;
  Matches  167;  Conservative    1;  Mismatches  179;  Indels    0;  Gaps    0;

Qy        104 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 163
              |||  |  ||   | |   | | |||     | | ||   ||||||  ||  || ||| |
Db        162 GGATGGGTGGGTGGATCCATGGGTGGGTGTGTGGGTGGATGGACAGATGGGTTGATTGGT 221

Qy        164 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 223
              | | | |  |  ||   | | | ||  |||   | |   ||   |||  |  ||   | |
Db        222 GGGTGAATGGATGGATGGTTGGATCCATGGGTAGGTGAGTGGATGGATTGGTGGATAGAT 281

Qy        224 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 283
              || | | | |  ||   | | | |||   | | | | | | | ||    | | | | | |
Db        282 GGATGGATGGATGGATGGATGGATGGATGGATAGATGGATGGGTGAGTGGATAGATGGAT 341

Qy        284 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 343
               | |||       | | | | | | |   | | | |   | | |||   | | | |   |
Db        342 GGATGGATGAGCAGATGGTTGGGTTGATGGGTGGGTGAATGGATGGATGGTTGGATCCAT 401

Qy        344 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403
               | | |     ||| | | | |:||    | | | | | | | |||   |   | | | |
Db        402 GGGTAGGTGAGTTGATGGATTGSTGAGTGGGTGGGTGGATGGATGGATGGACGGATGGAT 461

Qy        404 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAG 450
               | | |   | ||| |||| |||  || ||| || | | | | || |
Db        462 GGATAGATGGATTGGTGAGTGGATGGTTGGGTTGATGGATGGTTATG 508


RESULT 63
US-10-027-632-143383
; Sequence 143383, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 143383
;   LENGTH: 588
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-143383

  Query Match             6.0%;  Score 60.2;  DB 13;  Length 588;
  Best Local Similarity   48.1%;  Pred. No. 2.5e-07;
  Matches  167;  Conservative    1;  Mismatches  179;  Indels    0;  Gaps    0;

Qy        104 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 163
              |||  |  ||   | |   | | |||     | | ||   ||||||  ||  || ||| |
Db        162 GGATGGGTGGGTGGATCCATGGGTGGGTGTGTGGGTGGATGGACAGATGGGTTGATTGGT 221

Qy        164 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 223
              | | | |  |  ||   | | | ||  |||   | |   ||   |||  |  ||   | |
Db        222 GGGTGAATGGATGGATGGTTGGATCCATGGGTAGGTGAGTGGATGGATTGGTGGATAGAT 281

Qy        224 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 283
              || | | | |  ||   | | | |||   | | | | | | | ||    | | | | | |
Db        282 GGATGGATGGATGGATGGATGGATGGATGGATAGATGGATGGGTGAGTGGATAGATGGAT 341

Qy        284 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 343
               | |||       | | | | | | |   | | | |   | | |||   | | | |   |
Db        342 GGATGGATGAGCAGATGGTTGGGTTGATGGGTGGGTGAATGGATGGATGGTTGGATCCAT 401

Qy        344 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403
               | | |     ||| | | | |:||    | | | | | | | |||   |   | | | |
Db        402 GGGTAGGTGAGTTGATGGATTGSTGAGTGGGTGGGTGGATGGATGGATGGACGGATGGAT 461

Qy        404 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAG 450
               | | |   | ||| |||| |||  || ||| || | | | | || |
Db        462 GGATAGATGGATTGGTGAGTGGATGGTTGGGTTGATGGATGGTTATG 508


; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 143383
;   LENGTH: 588
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-143383

  Query Match             6.0%;  Score 60.2;  DB 13;  Length 588;
  Best Local Similarity   48.1%;  Pred. No. 2.5e-07;
  Matches  167;  Conservative    1;  Mismatches  179;  Indels    0;  Gaps    0;

Qy        104 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 163
              |||  |  ||   | |   | | |||     | | ||   ||||||  ||  || ||| |
Db        162 GGATGGGTGGGTGGATCCATGGGTGGGTGTGTGGGTGGATGGACAGATGGGTTGATTGGT 221

Qy        164 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 223
              | | | |  |  ||   | | | ||  |||   | |   ||   |||  |  ||   | |
Db        222 GGGTGAATGGATGGATGGTTGGATCCATGGGTAGGTGAGTGGATGGATTGGTGGATAGAT 281

Qy        224 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 283
              || | | | |  ||   | | | |||   | | | | | | | ||    | | | | | |
Db        282 GGATGGATGGATGGATGGATGGATGGATGGATAGATGGATGGGTGAGTGGATAGATGGAT 341

Qy        284 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 343
               | |||       | | | | | | |   | | | |   | | |||   | | | |   |
Db        342 GGATGGATGAGCAGATGGTTGGGTTGATGGGTGGGTGAATGGATGGATGGTTGGATCCAT 401

Qy        344 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403
               | | |     ||| | | | |:||    | | | | | | | |||   |   | | | |
Db        402 GGGTAGGTGAGTTGATGGATTGCTGAGTGGGTGGGTGGATGGATGGATGGACGGATGGAT 461

Qy        404 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAG 450
               | | |   | ||| |||| |||  |: ||| || | | | | || |
Db        462 GGATAGATGGATTGGTGAGTGGATGGWTGGGTTGATGGATGGTTATG 508


RESULT 64
US-10-027-632-183437
; Sequence 183437, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
```

```
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 183437
;   LENGTH: 553
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-183437

  Query Match             5.9%;  Score 59;  DB 13;  Length 553;
  Best Local Similarity   49.8%;  Pred. No. 5.4e-07;
  Matches  146;  Conservative    1;  Mismatches  146;  Indels    0;  Gaps    0;

Qy        127 TGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGC 186
              | | || |||| ||| || |          |  |  ||     |||||||   ||| || |
Db         48 TAGGCTTCTTGTTGAAGGAAGTAATCATTTAATGTCTCCTTAGACAGTATTTCTGATTTC 107

Qy        187 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGC 246
              |  |    | |  |     | |   |   |  :   ||     ||   | || |
Db        108 ACTTTAAATTTCAAGTGTCCTGAAAGTTCCCGAARTGAGCCATGCTGTTTTGTTGTTGTT 167

Qy        247 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 306
                |||  |  ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  |
Db        168 GTTCCTTTTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC 227

Qy        307 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 366
              |  | | ||| ||  ||| ||  |   ||  ||  ||| ||  |   ||  ||  ||| |
Db        228 TTTCTTTCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 287

Qy        367 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              |  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||
Db        288 TCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 340


RESULT 65
US-10-017-128-3/c
; Sequence 3, Application US/10017128
; Publication No. US20030124536A1
; GENERAL INFORMATION:
;  APPLICANT: McCarthy, Jeanette
;  TITLE OF INVENTION: DIAGNOSIS AND TREATMENT FOR VASCULAR DISEASE
;  FILE REFERENCE: MMI-001
;  CURRENT APPLICATION NUMBER: US/10/017,128
;  CURRENT FILING DATE:  2001-12-14
;  PRIOR APPLICATION NUMBER: US 60/306,941
;  PRIOR FILING DATE: 2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/315,572
;  PRIOR FILING DATE: 2001-08-28
;  PRIOR APPLICATION NUMBER: US 60/327,488
;  PRIOR FILING DATE: 2001-10-05
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 159095
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-017-128-3

  Query Match             5.9%;  Score 58.8;  DB 14;  Length 159095;
  Best Local Similarity   58.1%;  Pred. No. 1.4e-06;
  Matches  122;  Conservative    0;  Mismatches   87;  Indels    1;  Gaps    1;

Qy        220 TGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 279
              | ||  ||| ||        | ||  ||  | | ||| ||| ||  ||  |||   |
Db     104955 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCCTTTCTTTC 104896

Qy        280 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 339
              | ||| ||   || ||| ||| ||| ||  |   ||| ||| ||   | || | ||| ||
Db     104895 TTCTCTCT-CTCTTCTTTCTTTCTCTCTCTCCCTTCCTTCCTTCATTGCTTCCTTCCT 104837

Qy        340 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 399
              ||||  ||  | | ||| ||| || |||  | | ||| ||| || |||  | | ||| ||
Db     104836 TGCTTCCTTCCTTCCTTCCTTCCTTGCTTCCTTCCTTCCTTCCTTGCTTCCTTCCTTCCT 104777

Qy        400 TGCTCGCTGGCCTGCTTGCTGAGGGGACAG 429
              |  |  || || | ||| ||     |||||
Db     104776 TCTTTCCTTGCTTCCTTCCTTCCAAGACAG 104747


RESULT 66
US-10-262-552-33
; Sequence 33, Application US/10262552
; Publication No. US20030125289A1
; GENERAL INFORMATION:
;  APPLICANT: Gelb, Bruce D.
;  APPLICANT:  Tartaglia, Marco
;  TITLE OF INVENTION: NOONAN SYNDROME GENE
;  FILE REFERENCE: 2420/1J859-US1
;  CURRENT APPLICATION NUMBER: US/10/262,552
;  CURRENT FILING DATE:  2002-10-01
;  PRIOR APPLICATION NUMBER: US 60/326,532
;  PRIOR FILING DATE: 2001-10-01
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
;   LENGTH: 300000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(300000)
;   OTHER INFORMATION: where n may be a or g or c or t/u, unknown or other
US-10-262-552-33

  Query Match             5.9%;  Score 58.6;  DB 14;  Length 300000;
  Best Local Similarity   61.4%;  Pred. No. 1.7e-06;
  Matches   94;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        251 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 310
              || ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db      70489 CTTCTTTCTTTCTTTCTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCC 70548

Qy        311 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 370
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db      70549 TTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTTCTTTCTTTCTTTCTTTC 70608

Qy        371 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403
              | ||| ||| ||  ||  | | ||| ||| ||
Db      70609 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 70641


RESULT 67
US-10-027-632-170164
; Sequence 170164, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
```

```
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 170164
;   LENGTH: 489
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-170164

  Query Match             5.8%;  Score 58.2;  DB 13;  Length 489;
  Best Local Similarity   57.9%;  Pred. No. 8.8e-07;
  Matches   99;  Conservative    2;  Mismatches   70;  Indels    0;  Gaps    0;

Qy        246 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 305
              ||   ||  |  ||| ||  ||  | | ||| |||:||  || || | |||||| || :
Db        173 CTCTTCTCTTCTCTTTCTTTCTTTCTTTCTTTCTTKCTTTCTTGCTTTCTTGCTTTCTTK 232

Qy        306 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
              ||  | | ||| |||  ||  |    | ||| ||| ||  ||  | | ||| ||| ||
Db        233 CTTTCTTTCTTTCTTTTTCCTTCCTTTYCTTTCTTYCTTTCTTTCTTTCTTTCTTTCTTT 292

Qy        366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 416
              ||  | | ||| ||| ||  ||  | | ||| |||  |  ||    | |||
Db        293 CTTTCTTTCTTTCTTTCTTTCTTTCTTCCTTCCTTCTTTCCTTCTTTTCTT 343


RESULT 68
US-10-027-632-170165
; Sequence 170165, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 170165
;   LENGTH: 489
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-170165

  Query Match             5.8%;  Score 58.2;  DB 13;  Length 489;
  Best Local Similarity   57.9%;  Pred. No. 8.8e-07;
  Matches   99;  Conservative    2;  Mismatches   70;  Indels    0;  Gaps    0;

Qy        246 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 305
              ||   ||  |  ||| ||  ||  | | ||| |||:||  || || | |||||| || :
Db        173 CTCTTCTCTTCTCTTTCTTTCTTTCTTTCTTTCTTKCTTTCTTGCTTTCTTGCTTTCTTK 232

Qy        306 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
              ||  | | ||| |||  ||  |    | ||| ||| ||  ||  | | ||| ||| ||
Db        233 CTTTCTTTCTTTCTTTTTCCTTCCTTTYCTTTCTTYCTTTCTTTCTTTCTTTCTTTCTTT 292

Qy        366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 416
              ||  | | ||| ||| ||  ||  | | ||| |||  |  ||    | |||
Db        293 CTTTCTTTCTTTCTTTCTTTCTTTCTTCCTTCCTTCTTTCCTTCTTTTCTT 343


RESULT 69
US-10-027-632-170163
; Sequence 170163, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 170163
;   LENGTH: 978
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-170163

  Query Match             5.8%;  Score 58.2;  DB 13;  Length 978;
  Best Local Similarity   57.9%;  Pred. No. 9.7e-07;
  Matches   99;  Conservative    2;  Mismatches   70;  Indels    0;  Gaps    0;

Qy        246 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 305
              ||   ||  |  ||| ||  ||  | | ||| |||:||  || || | |||||| || :
Db        173 CTCTTCTCTTCTCTTTCTTTCTTTCTTTCTTTCTTKCTTTCTTGCTTTCTTGCTTTCTTK 232

Qy        306 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
              ||  | | ||| |||  ||  |    | ||| ||| ||  ||  | | ||| ||| ||
Db        233 CTTTCTTTCTTTCTTTTTCCTTCCTTTYCTTTCTTYCTTTCTTTCTTTCTTTCTTTCTTT 292

Qy        366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 416
              ||  | | ||| ||| ||  ||  | | ||| |||  |  ||    | |||
Db        293 CTTTCTTTCTTTCTTTCTTTCTTTCTTCCTTCCTTCTTTCCTTCTTTTCTT 343


RESULT 70
US-10-027-632-68339/c
; Sequence 68339, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
```

```
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68339
;    LENGTH: 439
;    TYPE: DNA
;    ORGANISM: Human
US-10-027-632-68339

  Query Match             5.8%;  Score 57.6;  DB 13;  Length 439;
  Best Local Similarity   58.2%;  Pred. No. 1.3e-06;
  Matches   99;  Conservative    1;  Mismatches   70;  Indels    0;  Gaps    0;

Qy        250 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 309
              | | ||| ||  ||   |    | ||| ||: ||  ||  | | ||| ||| ||  ||
Db        215 CTTTCTTTCTCTCTTTTTCTTTTTCTTTCTYTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 156

Qy        310 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 369
              | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||
Db        155 CTTCCTTCCTTCCTTCCTTCCTTCCTTYCTTCCTTCCTTCCTTCCTTTCTTTCTTTCTTT 96

Qy        370 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              | | ||| ||| ||  ||  | | ||| ||| |   ||  ||| ||| ||
Db         95 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCATTCTTTCCTTCTTTCT 46


RESULT 71
US-10-027-632-68340/c
; Sequence 68340, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68340
;    LENGTH: 439
;    TYPE: DNA
;    ORGANISM: Human
US-10-027-632-68340

  Query Match             5.8%;  Score 57.6;  DB 13;  Length 439;
  Best Local Similarity   58.2%;  Pred. No. 1.3e-06;
  Matches   99;  Conservative    1;  Mismatches   70;  Indels    0;  Gaps    0;

Qy        250 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 309
              | | ||| ||  ||   |    | ||| ||: ||  ||  | | ||| ||| ||  ||
Db        215 CTTTCTTTCTCTCTTTTTCTTTTTCTTTCTYTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 156

Qy        310 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 369
              | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||
Db        155 CTTCCTTCCTTCCTTCCTTCCTTCCTTYCTTCCTTCCTTCCTTCCTTTCTTTCTTTCTTT 96

Qy        370 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              | | ||| ||| ||  ||  | | ||| ||| |   ||  ||| ||| ||
Db         95 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCATTCTTTCCTTCTTTCT 46


RESULT 72
US-10-027-632-295071/c
; Sequence 295071, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 295071
;    LENGTH: 439
;    TYPE: DNA
;    ORGANISM: Human
US-10-027-632-295071

  Query Match             5.8%;  Score 57.6;  DB 13;  Length 439;
  Best Local Similarity   58.2%;  Pred. No. 1.3e-06;
  Matches   99;  Conservative    1;  Mismatches   70;  Indels    0;  Gaps    0;

Qy        250 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 309
              | | ||| ||  ||   |    | ||| ||: ||  ||  | | ||| ||| ||  ||
Db        215 CTTTCTTTCTCTCTTTTTCTTTTTCTTTCTYTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 156

Qy        310 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 369
              | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||
Db        155 CTTCCTTCCTTCCTTCCTTCCTTCCTTYCTTCCTTCCTTCCTTCCTTTCTTTCTTTCTTT 96

Qy        370 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              | | ||| ||| ||  ||  | | ||| ||| |   ||  ||| ||| ||
Db         95 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCATTCTTTCCTTCTTTCT 46


RESULT 73
US-10-027-632-295072/c
; Sequence 295072, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
```

```
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 295072
;   LENGTH: 439
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-295072

  Query Match             5.8%;  Score 57.6;  DB 13;  Length 439;
  Best Local Similarity   58.2%;  Pred. No. 1.3e-06;
  Matches   99;  Conservative    1;  Mismatches   70;  Indels    0;  Gaps    0;

Qy        250 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 309
              | | ||| ||  ||   |    | ||| ||: ||  ||  | | ||| ||| ||  ||
Db        215 CTTTCTTTCTCTCTTTTTCTTTTTCTTTCTYTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 156

Qy        310 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 369
              | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||
Db        155 CTTCCTTCCTTCCTTCCTTCCTTCCTTYCTTCCTTCCTTCCTTCCTTTCTTTCTTTCTTT 96

Qy        370 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              | |.||| ||| ||  ||  | | ||| ||| |   ||  ||| ||| ||
Db         95 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCATTCTTTCCTTCTTTCT 46


RESULT 74
US-10-027-632-37032/c
; Sequence 37032, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 37032
;   LENGTH: 549
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-37032

  Query Match             5.8%;  Score 57.6;  DB 13;  Length 549;
  Best Local Similarity   58.2%;  Pred. No. 1.3e-06;
  Matches   99;  Conservative    1;  Mismatches   70;  Indels    0;  Gaps    0;

Qy        250 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 309
              | | ||| ||  ||   |    | ||| ||: ||  ||  | | ||| ||| ||  ||
Db        215 CTTTCTTTCTCTCTTTTTCTTTTTCTTTCTYTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 156

Qy        310 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 369
              | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||
Db        155 CTTCCTTCCTTCCTTCCTTCCTTCCTTYCTTCCTTCCTTCCTTCCTTTCTTTCTTTCTTT 96

Qy        370 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              | | ||| ||| ||  ||  | | ||| ||| |   ||  ||| ||| ||
Db         95 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCATTCTTTCCTTCTTTCT 46


RESULT 75
US-10-027-632-37033/c
; Sequence 37033, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 37033
;   LENGTH: 549
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-37033

  Query Match             5.8%;  Score 57.6;  DB 13;  Length 549;
  Best Local Similarity   58.2%;  Pred. No. 1.3e-06;
  Matches   99;  Conservative    1;  Mismatches   70;  Indels    0;  Gaps    0;

Qy        250 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 309
              | | ||| ||  ||   |    | ||| ||: ||  ||  | | ||| ||| ||  ||
Db        215 CTTTCTTTCTCTCTTTTTCTTTTTCTTTCTYTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 156

Qy        310 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 369
              | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||
Db        155 CTTCCTTCCTTCCTTCCTTCCTTCCTTYCTTCCTTCCTTCCTTCCTTTCTTTCTTTCTTT 96

Qy        370 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              | | ||| ||| ||  ||  | | ||| ||| |   ||  ||| ||| ||
Db         95 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCATTCTTTCCTTCTTTCT 46


RESULT 76
US-09-803-719-1534/c
; Sequence 1534, Application US/09803719
; Publication No. US20030044783A1
; GENERAL INFORMATION:
;  APPLICANT: Williams, Lewis T.
;  APPLICANT:  Escobedo, Jaime
;  APPLICANT:  Innis, Michael A.
;  APPLICANT:  Garcia, Pablo Dominiguez
;  APPLICANT:  Sudduth-Klinger, Julie
;  APPLICANT:  Reinhard, Christoph
;  APPLICANT:  Giese, Klaus
;  APPLICANT:  Randazzo, Filippo
;  APPLICANT:  Kennedy, Giulia C.
;  APPLICANT:  Pot, David
;  APPLICANT:  Kassam, Altaf
;  APPLICANT:  Lamson, George
;  APPLICANT:  Drmanac, Radoje
;  APPLICANT:  Crkvenjakov, Radomir
;  APPLICANT:  Dickson, Mark
;  APPLICANT:  Drmanac, Snezana
;  APPLICANT:  Labat, Ivan
```

```
;  APPLICANT:  Leshkowitz, Dena
;  APPLICANT:  Kita, David
;  APPLICANT:  Garcia, Veronica
;  APPLICANT:  Jones, Lee William
;  APPLICANT:  Stache-Crain, Birgit
;  TITLE OF INVENTION: Human Genes and Gene Products
;  FILE REFERENCE: 1624.002
;  CURRENT APPLICATION NUMBER: US/09/803,719
;  CURRENT FILING DATE:  2001-03-09
;  PRIOR APPLICATION NUMBER: 60/188,609
;  PRIOR FILING DATE: 2000-03-09
;  NUMBER OF SEQ ID NOS: 2396
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1534
;    LENGTH: 412
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (1)...(412)
;    OTHER INFORMATION: n = A,T,C or G
US-09-803-719-1534

  Query Match             5.7%;  Score 56.6;  DB 11;  Length 412;
  Best Local Similarity   48.1%;  Pred. No. 2.4e-06;
  Matches  161;  Conservative    0;  Mismatches  174;  Indels    0;  Gaps    0;

Qy        391 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAG 450
              | ||| ||| |  | |||     | ||| | | ||      ||| |    | | |
Db        377 CGGCTGGCTCGTCCCCTGCAGGTCCTGCCGGGCGGGGGGCCTGGGGTCTCCGATGGGCTCT 318

Qy        451 GGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 510
              ||  ||  || |  |   | | ||   |  |||   ||||| |    |  |||||
Db        317 CGGCTGGCTCGTCCCCTGCAGGTCCTGCCGGGCGGGGGGCCTGGGGTCTCCGATGGGCTCT 258

Qy        511 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCC 570
              ||   |||  |         |  ||||  |   ||   ||||| |    |  ||| ||
Db        257 CGGCTGGCTCATCCCTTGTAGGTCCTGCCGGGCTGGGGCCTGGGGTCTCCGATGGGGTCT 198

Qy        571 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 630
              |||   ||| ||         |  ||||| |  |||  |||||| |    |  |||||
Db        197 TGGCTGGCTCGTCCCCTGCAGGTCCTGCTGGGCGGGGGCCTGGGGTCTCCGATGGGCTCT 138

Qy        631 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 690
              || | ||| ||         |  ||||| |  |||   ||||| |    |  |||||
Db        137 CGGCCGGCTCGTCCCCTGCAGGTCCTGCTGGGCGGGGGCCTGGGGTCTCCGATGGGCTCT 78

Qy        691 CTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCG 725
              | |||||| |    ||      |  |  |||| |
Db         77 CGGCTGGCTCGTCCCCTGCAGGTCCTGCTGGGCTG 43


RESULT 77
US-09-835-232-7
; Sequence 7, Application US/09835232
; Patent No. US20020098489A1
; GENERAL INFORMATION:
;  APPLICANT: Leder, Philip
;  APPLICANT:  Leader, Benjamin
;  TITLE OF INVENTION: FORMIN-2 NUCLEIC ACIDS AND POLYPEPTIDES
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: 00383/052002
;  CURRENT APPLICATION NUMBER: US/09/835,232
;  CURRENT FILING DATE:  2001-04-12
;  PRIOR APPLICATION NUMBER: US 60/196,811
;  PRIOR FILING DATE: 2000-04-13
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;    LENGTH: 170834
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (1)...(170834)
;    OTHER INFORMATION: n= A,T,C, or G
US-09-835-232-7

  Query Match             5.6%;  Score 56.2;  DB 9;  Length 170834;
  Best Local Similarity   58.8%;  Pred. No. 7.4e-06;
  Matches   97;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy        239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
              ||||| |||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| |
Db      81486 CTGCTTGCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC 81545

Qy        299 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 358
              || ||  ||  | | ||   || ||  ||  | | ||| ||| ||  ||  | | ||| |
Db      81546 TTTCTTTCTTTCTTTCTCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC 81605

Qy        359 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403
              || |||  |  | | |||  || ||   |    | ||| ||| ||
Db      81606 TTTCTCTTTTTCTTTCTTTGTTTCTTTGTTTGTTTCTTTCTTTCT 81650


RESULT 78
US-09-864-761-9961/c
; Sequence 9961, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
```

```
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 9961
;   LENGTH: 573
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO Z82201.1
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
US-09-864-761-9961

  Query Match             5.6%;  Score 56;  DB 9;  Length 573;
  Best Local Similarity   51.6%;  Pred. No. 3.7e-06;
  Matches  128;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy        495 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 554
              ||||  ||| ||       | ||||  ||||||       | ||||  ||||||      |
Db        572 CTGCATGTACGGAATGATGCCTGCGTGTAGGGAATGATGCCTGCATGTAGGGAATGATGC 513

Qy        555 CTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGG 614
              ||||  ||||||       | |||||  ||||||       | |||   ||||||      |
Db        512 CTGCGTGTAGGGTATGATGCCTGCTTGTAGGGAATAATGCCTGTGTGTAGGGAATGATGC 453

Qy        615 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 674
              |||   ||| ||      | ||||  |||  |       | ||||  ||||||      |
Db        452 CTGTGTGTATGGAATGATGCCTGCATGTACAGAATAATGCCTGCACGTAGGGAATGATGA 393

Qy        675 CTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGC 734
              | ||  ||||||     |||   | ||  ||      |||| ||| |   |     |||
Db        392 CCGCGTGTAGGGTATGATGCCTACATGTATGGAATGATGCTTGTGTGTGTGGAGTACTGA 333

Qy        735 TACTAGGG 742
              | ||  ||
Db        332 TGCTGTGG 325


RESULT 79
US-09-864-761-26463/c
; Sequence 26463, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 26463
;   LENGTH: 573
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO Z82201.1
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
;   OTHER INFORMATION: SWISSPROT HIT: P04929, EVALUE 5.00e-02
;   OTHER INFORMATION: NT HIT: gi11426388, EVALUE 2.30e-02
;   OTHER INFORMATION: EST_HUMAN HIT: BE407615.1, EVALUE 3.50e-02
US-09-864-761-26463

  Query Match             5.6%;  Score 56;  DB 9;  Length 573;
  Best Local Similarity   51.6%;  Pred. No. 3.7e-06;
  Matches  128;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy        495 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 554
              ||||  ||| ||       | ||||  ||||||       | ||||  ||||||      |
Db        572 CTGCATGTACGGAATGATGCCTGCGTGTAGGGAATGATGCCTGCATGTAGGGAATGATGC 513

Qy        555 CTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGG 614
              ||||  ||||||       | |||||  ||||||       | |||   ||||||      |
Db        512 CTGCGTGTAGGGTATGATGCCTGCTTGTAGGGAATAATGCCTGTGTGTAGGGAATGATGC 453

Qy        615 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 674
              |||   ||| ||      | ||||  |||  |       | ||||  ||||||      |
Db        452 CTGTGTGTATGGAATGATGCCTGCATGTACAGAATAATGCCTGCACGTAGGGAATGATGA 393

Qy        675 CTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGC 734
              | ||  ||||||     |||   | ||  ||      |||| ||| |   |     |||
Db        392 CCGCGTGTAGGGTATGATGCCTACATGTATGGAATGATGCTTGTGTGTGTGGAGTACTGA 333

Qy        735 TACTAGGG 742
              | ||  ||
Db        332 TGCTGTGG 325


RESULT 80
US-10-027-632-270367/c
; Sequence 270367, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
```

```
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 270367
;   LENGTH: 502
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-270367

  Query Match             5.5%;  Score 55.2;  DB 13;  Length 502;
  Best Local Similarity   54.4%;  Pred. No. 6.1e-06;
  Matches  111;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy        254 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 313
              |||  || ||  ||  | | |||  || ||| ||  | | ||| ||| ||  ||  | |
Db        306 CTTTATTTCTTTCTTTCTTTCTTTTTTTCTCTCTTTCTTTCTTTCTTTCTTTCTYTCTTT 247

Qy        314 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 373
              ||| ||| ||  ||  | | ||| ||| ||  ||  ||| ||| |||  |  ||| | |
Db        246 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCCTTCTTTCTTCTTTCCTGCCTTA 187

Qy        374 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGG 433
              ||  | | ||  ||  | |  || ||  ||  ||    | |||  |          | ||
Db        186 CTGCCATCCTGCCTTTCTTCTTTCCTCCCTTCCTCCTTTCCTTCTTTCATCCTGCCTTGG 127

Qy        434 GCTGCTTGCTTGCTAAGGGGACGG 457
               ||||     ||||  |  ||| |
Db        126 CCTGCCAAAGTGCTGGGATGACTG 103


RESULT 81
US-10-027-632-270368/c
; Sequence 270368, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 270368
;   LENGTH: 502
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-270368

  Query Match             5.5%;  Score 55.2;  DB 13;  Length 502;
  Best Local Similarity   54.4%;  Pred. No. 6.1e-06;
  Matches  111;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy        254 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 313
              |||  || ||  ||  | | |||  || ||| ||  | | ||| ||| ||  ||  | |
Db        306 CTTTATTTCTTTCTTTCTTTCTTTTTTTCTCYCTTTCTTTCTTTCTTTCTTTCTCTCTTT 247

Qy        314 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 373
              ||| ||| ||  ||  | | ||| ||| ||  ||  ||| ||| |||  |  ||| | |
Db        246 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCCTTCTTTCTTCTTTCCTGCCTTA 187

Qy        374 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGG 433
              ||  | | ||  ||  | |  || ||  ||  ||    | |||  |          | ||
Db        186 CTGCCATCCTGCCTTTCTTCTTTCCTCCCTTCCTCCTTTCCTTCTTTCATCCTGCCTTGG 127

Qy        434 GCTGCTTGCTTGCTAAGGGGACGG 457
               ||||     ||||  |  ||| |
Db        126 CCTGCCAAAGTGCTGGGATGACTG 103


RESULT 82
US-10-027-632-174148
; Sequence 174148, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 174148
;   LENGTH: 631
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-174148

  Query Match             5.5%;  Score 55.2;  DB 13;  Length 631;
  Best Local Similarity   48.4%;  Pred. No. 6.3e-06;
  Matches  183;  Conservative    1;  Mismatches  189;  Indels    5;  Gaps    1;

Qy        451 GGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 510
              |||   |  | || |   | ||    ||  || | | ||   || || |    |||  ||
Db          9 GGGGGTGCTCCCCGGGGCGCTGTGCCGACGGTGGTGGTGGGAGGGTGTTCCCGGGGGCGC 68

Qy        511 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCC 570
              || ||     || | |  |   || || |    ||| ||||| || |   || | |
Db         69 TGTGCCAACGGTGGTGGGGGAAGGGTGTTCCCGGGGGTGCTGTGCCGACGGTGGTGGGGG 128

Qy        571 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 630
               || ||||    ||| ||||| || | |  | | |      || ||||    ||| |||
Db        129 GAGGGTGCTCCCCGGGGTGCTGTGCCGATGTTGGTGGGGGGAGGGTGCTCCCTGGGGTGC 188

Qy        631 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG-----CTGCTAGTAGG 685
              || || | | || | | ||   || || |    |||      |      | |      ||
Db        189 TGTGCCGATGGTGGTGGTGGGAGGGTGTTCCCCGGGGGCGCTGTGCCGACGGTGGTGGGG 248

Qy        686 GCTGCCTGCTGGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTG 745
              |  |  |||| || |  |||| |   | | :||| |  |   || |||| |  ||| ||
```

```
Db        249 GAAGGGTGCTCCCTGGGGTGCTGGGCAGATGRTGGTGGGGGGAGCTTGCTCCCTGGGGTG 308

Qy        746 CTGTGCAGCTGGGAGAACAGAGTAGGGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGG 805
              ||| || | |||  |    | |  ||    ||     | |||     |   ||    ||
Db        309 CTGGGCCGATGGTGGTGGGGGGAGGGTGCTCCCCGRGGGTGCTGTGCTGATGGTGGTGGG 368

Qy        806 GCCAGACGAGGACACGGG 823
              ||  |  | |  | | ||
Db        369 GCTGGTGGTGCTCCCCGG 386


RESULT 83
US-10-027-632-174149
; Sequence 174149, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 174149
;   LENGTH: 631
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-174149

  Query Match             5.5%;  Score 55.2;  DB 13;  Length 631;
  Best Local Similarity   48.4%;  Pred. No. 6.3e-06;
  Matches  183;  Conservative    1;  Mismatches  189;  Indels    5;  Gaps    1;

Qy        451 GGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 510
              |||   |  | ||  |   | ||    ||  || | | ||   || || |    |||  ||
Db          9 GGGGGTGCTCCCCGGGGCGCTGTGCCGACGGTGGTGGTGGGAGGGTGTTCCCGGGGGCGC 68

Qy        511 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCC 570
              || ||     || | |  |   || || |    ||| ||||| || |   || | |
Db         69 TGTGCCAACGGTGGTGGGGGAAGGGTGTTCCCGGGGGTGCTGTGCCGACGGTGGTGGGGG 128

Qy        571 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 630
                || ||||    ||| ||||| || | |  | | |       || ||||    ||| |||
Db        129 GAGGGTGCTCCCCGGGGTGCTGTGCCGATGTTGGTGGGGGGAGGGTGCTCCCTGGGGTGC 188

Qy        631 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG-----CTGCTAGTAGG 685
              || || | | || | | ||  || || |    |||       |      | |      ||
Db        189 TGTGCCGATGGTGGTGGTGGGAGGGTGTTCCCCGGGGGCGCTGTGCCGACGGTGGTGGGG 248

Qy        686 GCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTG 745
              |  |  |||| || |  |||| |   | | :||| |  |   || |||| |  ||| ||
Db        249 GAAGGGTGCTCCCTGGGGTGCTGGGCAGATGRTGGTGGGGGGAGCTTGCTCCCTGGGGTG 308

Qy        746 CTGTGCAGCTGGGAGAACAGAGTAGGGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGG 805
              ||| || | |||  |    | |  ||    ||     | |||     |   ||    ||
Db        309 CTGGGCCGATGGTGGTGGGGGGAGGGTGCTCCCCGRGGGTGCTGTGCTGATGGTGGTGGG 368

Qy        806 GCCAGACGAGGACACGGG 823
              ||  |  | |  | | ||
Db        369 GCTGGTGGTGCTCCCCGG 386


RESULT 84
US-09-728-446-190/c
; Sequence 190, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668Alel Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 190
;   LENGTH: 345
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(345)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-190

  Query Match             5.5%;  Score 54.8;  DB 9;  Length 345;
  Best Local Similarity   50.7%;  Pred. No. 7.4e-06;
  Matches  110;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy        197 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 256
              || ||  | |||| |   |  |  |  |||          ||||||   |   ||||||
Db        235 GCCTGGGGCGGGGGCTACTGCTCCCCCTGCTGTTGGTCCCTGCTGCTGCATTTGCTGCTT 176

Qy        257 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 316
                  | |||||   ||| | | ||| | | || | | ||| | | |||||  ||||
Db        175 TTGCTATTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTG 116

Qy        317 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 376
              | | ||| | ||||  | ||| |   |  ||||   ||  | |  || | | ||    |
Db        115 GATGGCTGGATGGCTGGATTGGTNNATGNCTGGATGGCNGGATGNCTNGTTTGCTGNATG 56

Qy        377 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 413
                |   | |||    | |     ||| | ||  |||
Db         55 NNTGNATGNCTGNGNCGNTGCANAGCTGGATGCGCTG 19


RESULT 85
US-10-311-455-689/c
; Sequence 689, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Det
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 689
;   LENGTH: 12007
```

```
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-689

  Query Match             5.5%;  Score 54.8;  DB 12;  Length 12007;
  Best Local Similarity   57.6%;  Pred. No. 1.2e-05;
  Matches   98;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        242 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 301
              ||| ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| |||
Db       3049 CTCTCTTTCTTCCTTTCTTACTTCCTTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 2990

Qy        302 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 361
              ||  ||  | | ||| ||| ||  ||  | | | |   | |   |   ||| ||| |||
Db       2989 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCCTTTCTTCCTTCCTTCCTTTCTTTCTT 2930

Qy        362 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              ||  ||  | | ||| ||| |||  |    | ||| ||| ||  ||  ||
Db       2929 CTTTCTTCCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTACTTACC 2880


RESULT 86
US-10-175-523-75
; Sequence 75, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
;  APPLICANT: Brockman, Jeffrey
;  APPLICANT:  Evans, David
;  APPLICANT:  Hook, Derek
;  APPLICANT:  Klimczak, Leszek
;  APPLICANT:  Laeng, Pascal
;  APPLICANT:  Palfreyman, Michael
;  APPLICANT:  Rajan, Prithi
;  TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
;  FILE REFERENCE: 3235/1J795-US3
;  CURRENT APPLICATION NUMBER: US/10/175,523
;  CURRENT FILING DATE:  2002-06-18
;  PRIOR APPLICATION NUMBER: US 60/299,151
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: US 60/317,828
;  PRIOR FILING DATE: 2001-09-07
;  PRIOR APPLICATION NUMBER: US 60/325,150
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 60/333,047
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: US 60/349,936
;  PRIOR FILING DATE: 2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/361,834
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 197
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 75
;   LENGTH: 85548
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-175-523-75

  Query Match             5.5%;  Score 54.8;  DB 14;  Length 85548;
  Best Local Similarity   59.9%;  Pred. No. 1.6e-05;
  Matches  109;  Conservative    0;  Mismatches   72;  Indels    1;  Gaps    1;

Qy        239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
              || |||  |  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| |
Db      10084 CTTCTCTTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCC 10143

Qy        299 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG-CTCGCTGGCCTGCTTG 357
              ||  ||  ||  | | ||| ||| ||  ||  ||| ||| |||  ||| ||  |   ||
Db      10144 TTCCTTCCTTCCTTCCTTCCTTCCTTCCCTCCTTCCTTCCCTCCCTTCCTCTCTCT 10203

Qy        358 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 417
              ||  ||| ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | |||

Db      10204 CTCCCTCCCTCCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 10263

Qy        418 CT 419
              ||
Db      10264 CT 10265


RESULT 87
US-10-027-632-179264
; Sequence 179264, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 179264
;   LENGTH: 1223197
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1223197)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-179264

  Query Match             5.5%;  Score 54.8;  DB 13;  Length 1223197;
  Best Local Similarity   33.8%;  Pred. No. 2.4e-05;
  Matches   71;  Conservative   56;  Mismatches   83;  Indels    0;  Gaps    0;

Qy        207 GGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGC 266
              ||| |  ||  |||| | ::|     | ||||:   | :|| ::|||::   :   | :
Db    1203958 GGGGCYYTAAAGCTGGGGCYYTAAAGCTGGGGCYYTGCYYTGCYYTGCYYTGCYYTGCYY 1204017

Qy        267 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 326
              || ::|||::   :   | :|| ::|||::   :   | :|| ::|||::   :   | :
Db    1204018 TGCYYTGCYYTGCYYTGCYYTGCYYTGCYYTGCYYTGCYYTGCYYTGCYYTGCYYTGCYY 1204077

Qy        327 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 386
              || ::|||::   :   | :|| ::|||::   :   | :|| ::|||::   :   | :
Db    1204078 TGCYYTGCYYTGCYYTGCYYTGCYYTGCYYTGCYYTGCYYTGCYYTGCYYTGCYYTGCYY 1204137

Qy        387 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 416
              || ::|||::   :   | :|| ::|||::
Db    1204138 TGCYYTGCYYTGCYYTGCYYTGCYYTGCYY 1204167


RESULT 88
US-09-728-446-435/c
; Sequence 435, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
```

```
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 435
;   LENGTH: 354
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(354)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-435

  Query Match             5.5%;  Score 54.6;  DB 9;  Length 354;
  Best Local Similarity   55.2%;  Pred. No. 8.5e-06;
  Matches   96;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
              | ||||  | |  |     ||||   |   |||||   |  | ||  | | ||| | | |
Db        246 CCTGCTGTTGGTCCCTGCTGCTGCATTTGCTGCTTTTGCTATTGCTGGATGGCTGGATGG 187

Qy        286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
              ||||  |||| | | ||| | ||||  | |||||  | |  |||||   ||| | ||| | |
Db        186 CTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGGTGG 127

Qy        346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 399
               ||||  | | ||| | | | |||  | ||   | | |   |||     |   |
Db        126 ATGGCTGGATGGCTGGATGGNTGGTTTACTNAATGGNTNTATGGNTNNATNCNT 73


RESULT 89
US-10-027-632-141461/c
; Sequence 141461, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 141461
;   LENGTH: 707
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-141461

  Query Match             5.5%;  Score 54.6;  DB 13;  Length 707;
  Best Local Similarity   58.2%;  Pred. No. 9.4e-06;
  Matches   96;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy        211 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 270
              ||  | |||||  || |||   || ||| ||    ||  | |·||| ||| ||  ||  |
Db        168 CACCATGGCTGGAGGATTGAGGGTGGGGATGGGATCTTTCTTTCTTTCTTTCTTTCTTTC 109

Qy        271 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 330
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||| ||  |
Db        108 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTCTCTC 49

Qy        331 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 375
                ||  ||  ||| ||  |   ||| ||| ||| ||  |   ||
Db         48 TCTCTCTCTCTCTCTCTCTCTCTTTCTTTCTCTCTCTCTCTCT 4


RESULT 90
US-10-027-632-86169/c
; Sequence 86169, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 86169
;   LENGTH: 589
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-86169

  Query Match             5.4%;  Score 54;  DB 13;  Length 589;
  Best Local Similarity   56.9%;  Pred. No. 1.3e-05;
  Matches   99;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy        243 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 302
              || ||  |   ||| ||  ||| ||  | | ||| ||| ||  ||  | | ||| ||| |
Db        365 TCTCTCTCTCTCTTTCTGTCTCTCTCCCTTCCTTCCTTCCTTCCTTCCTTCCTTTC 306

Qy        303 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 362
              |  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| |
Db        305 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTCCTTCC 246

Qy        363 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 416
              |  ||  | | ||  ||  ||| ||  |    |  ||   || ||  | | |||
Db        245 TTCCTTCCATCCTCCCTCCCTCCCTCCCTCCATACCTCTTTCTCTCTCTTTCTT 192


RESULT 91
US-10-175-523-57/c
; Sequence 57, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
;  APPLICANT: Brockman, Jeffrey
;  APPLICANT:  Evans, David
;  APPLICANT:  Hook, Derek
;  APPLICANT:  Klimczak, Leszek
;  APPLICANT:  Laeng, Pascal
;  APPLICANT:  Palfreyman, Michael
;  APPLICANT:  Rajan, Prithi
;  TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
```

```
;  FILE REFERENCE: 3235/1J795-US3
;  CURRENT APPLICATION NUMBER: US/10/175,523
;  CURRENT FILING DATE:  2002-06-18
;  PRIOR APPLICATION NUMBER: US 60/299,151
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: US 60/317,828
;  PRIOR FILING DATE: 2001-09-07
;  PRIOR APPLICATION NUMBER: US 60/325,150
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 60/333,047
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: US 60/349,936
;  PRIOR FILING DATE: 2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/361,834
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 197
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57
;   LENGTH: 225883
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-175-523-57

  Query Match             5.4%;  Score 54;  DB 14;  Length 225883;
  Best Local Similarity   53.9%;  Pred. No. 3.2e-05;
  Matches  111;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy        240 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 299
              | ||  ||  | .| ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||
Db     194415 TCCTTCCTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 194356

Qy        300 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 359
              | ||  ||  | | ||  ||  ||| ||  |   ||  ||  ||| ||  |   ||  ||
Db     194355 TCCTTCCTTCCTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 194296

Qy        360 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
                ||| ||      ||| ||   || ||  | | ||| ||| |  |      || |   |
Db     194295 CTCTCTCTCTTTCCCTTTCTCTTTCTCTTTCTTTCTTCCCTGACCCAGTGATCATT 194236

Qy        420 GAGGGGACAGTAGGGCTGCTTGCTTG 445
              |||  || ||| |  | | || | ||
Db     194235 GAGTAGAAAGTTGTTCAGTTTCCATG 194210


RESULT 92
US-09-791-254-4/c
; Sequence 4, Application US/09791254
; Publication No. US20030148935A1
; GENERAL INFORMATION:
;  APPLICANT: Peymann, John A.
;  APPLICANT:  Carulli, John
;  APPLICANT:  Browning, Jeffrey
;  TITLE OF INVENTION: No. US20030148935A1el Nuclear Factor Polypeptides and Nucleic Aci
;  TITLE OF INVENTION:  Encoding Same
;  FILE REFERENCE: 15966-652 (Cura-152) US
;  CURRENT APPLICATION NUMBER: US/09/791,254
;  CURRENT FILING DATE:  2001-09-13
;  PRIOR APPLICATION NUMBER: USSN 60/184,241
;  PRIOR FILING DATE: 2000-02-23
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 2201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-791-254-4

  Query Match             5.4%;  Score 53.8;  DB 12;  Length 2201;
  Best Local Similarity   46.1%;  Pred. No. 1.8e-05;
  Matches  181;  Conservative    0;  Mismatches  212;  Indels    0;  Gaps    0;

Qy        333 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 392
              ||| |||    |  ||     | |||  |   |  | |  |   |  |    | | |     |
Db       1795 GCTGGCTCATCCCTTGTAGGTCCTGCCGGGCTGGGGCCTGGGGTCTCCGATGGGCTCTCG 1736

Qy        393 GCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGG 452
              ||| ||| |  | |||     | ||| | | ||      ||| |    | | |      |
Db       1735 GCTGGCTCGTCCCCTGCAGGTCCTGCCGGGCGGGGGCCTGGGGTCTCCGATGGGCTCTCG 1676

Qy        453 GACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 512
              |  ||  || |  |   | | | || | |  |||   ||||| |    | |||||   |
Db       1675 GCTGGCTCGTCCCCTGCAGGTCCTGCTGGGCGGGGGCCTGGGGTCTCCGATGGGCTCTCG 1616

Qy        513 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTG 572
              |   |||  |          |  ||||  |   ||   ||||| |    | ||| ||  |
Db       1615 GCTGGCTCATCCCTTGTAGGTCCTGCCGGGCTGGGGCCTGGGGTCTCCGATGGGGTCTCG 1556

Qy        573 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTG 632
              |   ||| ||         |  ||||| |  |||  ||||||  |    |  |||||   |
Db       1555 GCTGGCTCGTCCCCTGCAGGTCCTGCTGGGCGGGGGCCTGGGGTCTCCGATGGGCTCTCG 1496

Qy        633 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCT 692
              |   ||| ||          |  ||||| |   ||   ||||| |    | ||||| |
Db       1495 GCTGGCTCGTCCCCTGCAGGTCCTGCTGGGCTGGGGCCTGGGGTCTCCGATGGGCTCTCG 1436

Qy        693 GCTGGCTTGCTTGCTTGCTAGTGGGGCCG 725
              ||||||| |    ||      |  |  |||| |
Db       1435 GCTGGCTCGTCCCCTGCAGGTCCTGCTGGGCTG 1403


RESULT 93
US-09-791-254-1/c
; Sequence 1, Application US/09791254
; Publication No. US20030148935A1
; GENERAL INFORMATION:
;  APPLICANT: Peymann, John A.
;  APPLICANT:  Carulli, John
;  APPLICANT:  Browning, Jeffrey
;  TITLE OF INVENTION: No. US20030148935A1el Nuclear Factor Polypeptides and Nucleic
;  TITLE OF INVENTION:  Encoding Same
;  FILE REFERENCE: 15966-652 (Cura-152) US
;  CURRENT APPLICATION NUMBER: US/09/791,254
;  CURRENT FILING DATE:  2001-09-13
;  PRIOR APPLICATION NUMBER: USSN 60/184,241
;  PRIOR FILING DATE: 2000-02-23
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4065
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(3726)
US-09-791-254-1

  Query Match             5.4%;  Score 53.8;  DB 12;  Length 4065;
  Best Local Similarity   46.1%;  Pred. No. 2e-05;
  Matches  181;  Conservative    0;  Mismatches  212;  Indels    0;  Gaps    0;

Qy        333 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 392
              ||| |||    |  ||     | |||  |   |  | |  |   |  |    | | |     |
Db       3659 GCTGGCTCATCCCTTGTAGGTCCTGCCGGGCTGGGGCCTGGGGTCTCCGATGGGCTCTCG 3600

Qy        393 GCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGG 452
              ||| ||| |  | |||     | ||| | | ||      ||| |    | | |      |
Db       3599 GCTGGCTCGTCCCCTGCAGGTCCTGCCGGGCGGGGGCCTGGGGTCTCCGATGGGCTCTCG 3540

Qy        453 GACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 512
              |  ||  || |  |   | | | ||   |  |||   ||||| |    | |||||   |
Db       3539 GCTGGCTCGTCCCCTGCAGGTCCTGCCGGGCGGGGGCCTGGGGTCTCCGATGGGCTCTCG 3480

Qy        513 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTG 572
              |   |||  |          |  ||||  |   ||   ||||| |    | ||||||  |
Db       3479 GCTGGCTCATCCCTTGTAGGTCCTGCCGGGCTGGGGCCTGGGGTCTCCGATGGGCTCTCG 3420
```

```
Qy      "   573 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTG 632
Db       3419 GCTGGCTCGTCCCCTGCAGGTCCTGCCGGGCGGGGGCCTGGGGTCTCCGATGGGCTCTCG 3360

Qy        633 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCT 692
Db       3359 GCTGGCTCGTCCCCTGCAGGTCCTGCTGGGCGGGGGCCTGGGGTCTCCGATGGGCTCTCG 3300

Qy        693 GCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCG 725
Db       3299 GCTGGCTCGTCCCCTGCAGGTCCTGCTGGGCTG 3267


RESULT 94
US-10-027-632-179264/c
; Sequence 179264, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 179264
;   LENGTH: 1223197
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1223197)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-179264

  Query Match             5.4%;  Score 53.8;  DB 13;  Length 1223197;
  Best Local Similarity   56.5%;  Pred. No. 4.6e-05;
  Matches  135;  Conservative    2;  Mismatches   99;  Indels    3;  Gaps    2;

Qy        178 CCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGG 237
Db     471226 CATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATKCATGCATGCA-T 471168

Qy        238 GCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC--CTGCT 295
Db     471167 GATGCATGCATGCATGCATKCATGCATGCATGCATGCATGCATGAAATCAGCCAACTTAA 471108

Qy        296 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 355
Db     471107 GCCAACTTAGCTGGCTGGCTGGATGGCTGGATGGCTGGCTGGCTTGCTTACATTACTTAC 471048

Qy        356 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 414
Db     471047 TTATTGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGC 470989


RESULT 95
US-09-860-670-161
; Sequence 161, Application US/09860670
; Patent No. US20020165137A1
; GENERAL INFORMATION:
;  APPLICANT: Ruben et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA127P1
;  CURRENT APPLICATION NUMBER: US/09/860,670
;  CURRENT FILING DATE:  2001-05-21
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 289
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 161
;   LENGTH: 26555
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-860-670-161

  Query Match             5.4%;  Score 53.6;  DB 10;  Length 26555;
  Best Local Similarity   51.4%;  Pred. No. 3e-05;
  Matches  149;  Conservative    0;  Mismatches  139;  Indels    2;  Gaps    1;

Qy        132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191
Db       1395 TCCTTGACAAGGAGTCACTTGTCCAGCTTGTTTAGGATCCTTTGATGATGACTTCATGCC 1454

Qy        192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251
Db       1455 CCCAGGCCAGCCCTGGCCATCTCTTTTCTTTTCTTTTCTTTTCTTTCTTTCTTTCTTTCT 1514

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
Db       1515 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCCTTCTTTCT 1574

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG--CTTGCTTGCTCGCTGG 369
Db       1575 TTCTTTCTTCCTTTCTTTCTCTTTCTCTCTCTCTCTTTCTTTCTCTTTCTTTCTTTCTTT 1634

Qy        370 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
Db       1635 CCTTCTTTCTTTCTTTCTTCCTTTCTTTCTCTTTCTCTCTCTCTTTCT 1684


RESULT 96
US-10-007-010-10/c
; Sequence 10, Application US/10007010
; Publication No. US20030125275A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander H. Borchers
;  APPLICANT:  Kenneth W. Dobie
;  TITLE OF INVENTION: ANTISENSE MODULATION OF HCK EXPRESSION
;  FILE REFERENCE: RTS-0345
;  CURRENT APPLICATION NUMBER: US/10/007,010
;  CURRENT FILING DATE:  2001-12-04
;  NUMBER OF SEQ ID NOS: 87
; SEQ ID NO 10
;   LENGTH: 30000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (4334)...(4437)
;   OTHER INFORMATION: exon 5
;   NAME/KEY: exon:intron junction
;   LOCATION: (4437)...(4438)
;   OTHER INFORMATION: exon 5:intron 5
;   NAME/KEY: intron
;   LOCATION: (4438)...(8453)
;   OTHER INFORMATION: intron 5
;   NAME/KEY: intron:exon junction
;   LOCATION: (8453)...(8454)
;   OTHER INFORMATION: intron 5:exon 6
;   NAME/KEY: exon
;   LOCATION: (8454)...(8603)
;   OTHER INFORMATION: exon 6
```

```
;   NAME/KEY: exon:intron junction
;   LOCATION: (8603)...(8604)
;   OTHER INFORMATION: exon 6:intron 6
;   NAME/KEY: intron:exon junction
;   LOCATION: (8950)...(8951)
;   OTHER INFORMATION: intron 6:exon 7
;   NAME/KEY: exon
;   LOCATION: (8951)...(9103)
;   OTHER INFORMATION: exon 7
;   NAME/KEY: exon:intron junction
;   LOCATION: (9103)...(9104)
;   OTHER INFORMATION: exon 7:intron 7
;   NAME/KEY: intron
;   LOCATION: (9104)...(11187)
;   OTHER INFORMATION: intron 7
;   NAME/KEY: intron:exon junction
;   LOCATION: (11187)...(11188)
;   OTHER INFORMATION: intron 7:exon 8
;   NAME/KEY: exon
;   LOCATION: (11188)...(11367)
;   OTHER INFORMATION: exon 8
;   NAME/KEY: exon:intron junction
;   LOCATION: (11367)...(11368)
;   OTHER INFORMATION: exon 8:intron 8
;   NAME/KEY: intron
;   LOCATION: (11368)...(13127)
;   OTHER INFORMATION: intron 8
;   NAME/KEY: intron:exon junction
;   LOCATION: (13127)...(13128)
;   OTHER INFORMATION: intron 8:exon 9
;   NAME/KEY: exon
;   LOCATION: (13128)...(13204)
;   OTHER INFORMATION: exon 9
;   NAME/KEY: exon:intron junction
;   LOCATION: (13204)...(13205)
;   OTHER INFORMATION: exon 9:intron 9
;   NAME/KEY: intron
;   LOCATION: (13205)...(18422)
;   OTHER INFORMATION: intron 9
;   NAME/KEY: intron:exon junction
;   LOCATION: (18422)...(18423)
;   OTHER INFORMATION: intron 9:exon 10
;   NAME/KEY: exon
;   LOCATION: (18423)...(18576)
;   OTHER INFORMATION: exon 10
;   NAME/KEY: exon:intron junction
;   LOCATION: (18576)...(18577)
;   OTHER INFORMATION: exon 10:intron 10
;   NAME/KEY: intron
;   LOCATION: (18577)...(23563)
;   OTHER INFORMATION: intron 10
;   NAME/KEY: intron:exon junction
;   LOCATION: (23563)...(23564)
;   OTHER INFORMATION: intron 10:exon 11
;   NAME/KEY: exon
;   LOCATION: (23564)...(23695)
;   OTHER INFORMATION: exon 11
;   NAME/KEY: exon:intron junction
;   LOCATION: (23695)...(23696)
;   OTHER INFORMATION: exon 11:intron 11
;   NAME/KEY: intron
;   LOCATION: (23696)...(25876)
;   OTHER INFORMATION: intron 11
;   NAME/KEY: intron:exon junction
;   LOCATION: (25876)...(25877)
;   OTHER INFORMATION: intron 11:exon 12
;   NAME/KEY: exon
;   LOCATION: (25877)...(26412)
;   OTHER INFORMATION: exon 12
US-10-007-010-10

  Query Match             5.4%;  Score 53.6;  DB 14;  Length 30000;
  Best Local Similarity   57.9%;  Pred. No. 3.1e-05;
  Matches   95;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy        249 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 308
              ||||  ||  ||  |  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||
Db      19426 GCCTCTTTCTTTCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT 19367

Qy        309 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 368
              | | ||| ||| ||  ||      ||  ||| ||  ||  ||| ||| ||| ||  ||
Db      19366 TCTTTCTTTCTTTCTTTCTCTTTCTCTCTCTTTCTTTCTTTCCTTCTTTCTTTCTTTCTT 19307

Qy        369 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 412
              | | ||| ||| ||  ||  | | ||| ||| ||  ||   ||
Db      19306 TCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTTCT 19263


RESULT 97
US-10-101-510-127
; Sequence 127, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
;  APPLICANT: WAN, JACKSON
;  APPLICANT:  WANG, YIXIN
;  TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
;  FILE REFERENCE: 15117.0012
;  CURRENT APPLICATION NUMBER: US/10/101,510
;  CURRENT FILING DATE:  2002-03-20
;  PRIOR APPLICATION NUMBER: 60/276,947
;  PRIOR FILING DATE: 2001-03-20
;  NUMBER OF SEQ ID NOS: 805
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 127
;   LENGTH: 136328
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-101-510-127

  Query Match             5.4%;  Score 53.6;  DB 12;  Length 136328;
  Best Local Similarity   51.4%;  Pred. No. 3.8e-05;
  Matches  149;  Conservative    0;  Mismatches  139;  Indels    2;  Gaps    1;

Qy        132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191
              | ||||   ||| | || | | || ||||| | |||   |  |     ||  | |  ||
Db      24832 TCCTTGACAAGGAGTCACTTGTCCAGCTTGTTTAGGATCCTTTGATGATGACTTCATGCC 24891

Qy        192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251
              |  ||  ||   ||  |        ||  |   ||  |        | ||  ||  |
Db      24892 CCCAGGCCAGCCCTGGCCATCTCTTTTCTTTTCTTTTCTTTTCTTTCTTTCTTTCTTTCT 24951

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
              |.||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| |   ||  |
Db      24952 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCCTTCTTTCT 25011

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG--CTTGCTTGCTCGCTGG 369
              | ||| ||| ||  ||  |    |  ||  ||| ||  | |   ||| ||| ||  ||
Db      25012 TTCTTTCTTCCTTTCTTTCTCTTTCTCTCTCTCTCTTTCTTTCTCTTTCTTTCTTTCTTT 25071

Qy        370 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              ||| ||| ||| ||  ||  | | ||| ||   || ||  |   ||| ||
Db      25072 CCTTCTTTCTTTCTTTCTTCCTTTCTTTCTCTTTCTCTCTCTCTCTTTCT 25121


RESULT 98
US-10-017-161-2377
; Sequence 2377, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
```

```
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2377
;   LENGTH: 1297
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1297)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1097)
US-10-017-161-2377

  Query Match             5.3%;  Score 53.4;  DB 12;  Length 1297;
  Best Local Similarity   44.3%;  Pred. No. 2.2e-05;
  Matches  261;  Conservative    0;  Mismatches  326;  Indels    2;  Gaps    1;

Qy        153 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 212
              || ||  |||       |   | | |  ||   ||  |   |  | | |||     ||
Db        115 GCATGTATGCACGTATGTATGCATGTATGTGGGCATGTATGTATACATGCATGTAGGCAT 174

Qy        213 GTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 272
              ||| |  | |  || ||   | | |  |||  |   |  ||| | | ||   ||  |  |
Db        175 GTATGTATACATGCATGTAGGCATGTATGCGTGTATGTATGCCTCCATGTATGCATGTGT 234

Qy        273 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 332
              |  | | ||   |   |  ||| ||  |   |    |  |       ||   ||  |  |
Db        235 GTATACATGTATGTATGTATGCATGTGTATGCATGTGTGTATGCATGTGTATGCATGTAT 294

Qy        333 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 392
              |  ||  |||  |   |  | | ||  |     || |  ||  ||| ||   ||  |  |
Db        295 GTATGTATGCAGGTGTGTATCCATGTGTTTATACTAGTATGTATGCATGTATGCATGTAT 354

Qy        393 GCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGG 452
               | ||  ||   ||  |  ||| ||   | | |   ||| |  ||  ||| ||   | |
Db        355 ACGTGTATGTATGCATGTATGCATGTATACGTGTATGTATGCATGTATGCATGTATACGT 414

Qy        453 GACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 512
              |   ||| || ||  ||   |  |    ||| |  |||  |  |||  |||    ||
Db        415 GTATGTATGCATGTATGCATGTATACGTGTATGTATGCATGTATGCATGTATACGTGTAT 474

Qy        513 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGT--AGGGCTCC 570
              |  |||  ||| |  ||       ||   ||| |  ||   |  | |  ||  | | ||
Db        475 GTATGCATGTATGCATGTATACGTGTATGTATGCATGTACGTATACATGTGCATGCCTGT 534

Qy        571 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 630
                ||   |||    ||   |  |||  ||| |  |    |  |||  ||  |  |||
Db        535 GTACATGTATGTATACATGTGTGCATGCATGTATGTATGTGTGTATGCATGTGTGTATGC 594

Qy        631 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 690
              |  ||   | | |  ||   |  |||  | | |  ||      ||   |||    ||
Db        595 ATGCATGTATGCATGCATGTATGTGTGCATGCATGTGTGTATACATGTGTGTATACGTGT 654

Qy        691 CTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTA 739
              |||    | ||  ||  ||  |||  ||| |   |  | | ||  | ||
Db        655 CTGTGTACATGTATGTATGTATGCACGTGTGTATGTATACATGTAAGTA 703


RESULT 99
US-10-027-632-30729
; Sequence 30729, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30729
;   LENGTH: 766
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(766)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-30729

  Query Match             5.3%;  Score 53.2;  DB 13;  Length 766;
  Best Local Similarity   57.6%;  Pred. No. 2.3e-05;
  Matches  114;  Conservative    0;  Mismatches   83;  Indels    1;  Gaps    1;

Qy        222 CTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 281
              ||  ||| ||        | ||  ||  | | ||| ||| || |||  ||| ||| |||
Db        357 CTCTCTTTCTTTCTTCCTTCCTTCCTTTCTTCCTTTCTTTCTTGCTTTCCTTCTTTCTTT 416

Qy        282 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 341
              ||  |   | | | | ||| ||  ||  | | ||| ||| ||  ||  | | ||| |||
Db        417 CTTTCCTTCTTTCCTTCTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 476

Qy        342 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 401
              ||   |  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| | |
Db        477 CTTCTTTCCTTTCTT-CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCCTT 535

Qy        402 CTCGCTGGCCTGCTTGCT 419
              ||  ||  ||| ||| ||
Db        536 CTTTCTTTCCTTCTTTCT 553


RESULT 100
US-10-027-632-30730
; Sequence 30730, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30730
```

```
;   LENGTH: 766
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(766)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-30730

  Query Match             5.3%;  Score 53.2;  DB 13;  Length 766;
  Best Local Similarity   57.6%;  Pred. No. 2.3e-05;
  Matches  114;  Conservative    0;  Mismatches   83;  Indels    1;  Gaps    1;

Qy        222 CTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 281
              ||  ||| ||        | || ||  | | ||| ||| || |||  ||| ||| |||
Db        357 CTCTCTTTCTTTCTTCCTTCCTTCCTTTCTTCCTTTCTTTCTTGCTTTCCTTCTTTCTTT 416

Qy        282 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 341
              ||  |   | | | | ||| ||  ||  | | ||| ||| ||  ||  | | ||| |||
Db        417 CTTTCCTTCTTTCCTTCTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 476

Qy        342 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 401
              ||   |  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| | |
Db        477 CTTCTTTCCTTTCTT-CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCCTT 535

Qy        402 CTCGCTGGCCTGCTTGCT 419
              ||  ||  ||| ||| ||
Db        536 CTTTCTTTCCTTCTTTCT 553


Search completed: August 22, 2003, 00:50:14
Job time : 279.377 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 21, 2003, 20:57:58 ; Search time 2072.46 Seconds
(without alignments)
11727.335 Million cell updates/sec

Title: US-09-831-000-1_COPY_1_1000
Perfect score: 1000
Sequence: 1 gatcgggaaaacgcgagggg.....gtgcacggccaagttacaag 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	196.2	19.6	656	28	AZ282406	AZ282406 RPCI-23-1
	2	195.2	19.5	759	28	AZ646178	AZ646178 1M0512A15
c	3	165.6	16.6	821	29	BZ249369	BZ249369 CH230-409
c	4	148.2	14.8	684	29	BZ515928	BZ515928 BOMRD40TR
	5	145.6	14.6	545	28	AZ836738	AZ836738 2M0131P07
	6	142.6	14.3	799	14	CB499552	CB499552 ssalsrkc0
c	7	142.2	14.2	770	28	BZ182690	BZ182690 CH230-434
	8	141.4	14.1	540	28	BH276950	BH276950 CH230-48M
	9	140.6	14.1	749	28	BH484177	BH484177 BOGHS47TR
	10	140	14.0	712	29	BZ519739	BZ519739 BOMRL03TR
c	11	136.2	13.6	845	28	BH541157	BH541157 BOGLR58TR
	12	135.4	13.5	697	28	BH521799	BH521799 BOGXL17TF
	13	133.6	13.4	765	29	BZ221738	BZ221738 CH230-372
c	14	132.4	13.2	481	28	AZ828483	AZ828483 2M0105A06
	15	132	13.2	597	28	BH537125	BH537125 BOHCS81TR
c	16	132	13.2	1053	13	BX333427	BX333427 BX333427
c	17	130.8	13.1	894	29	CNS03LWH	AL250010.Tetraodon
c	18	129.8	13.0	640	29	DR37P5T	AL975317 Danio rer
	19	129.4	12.9	743	29	CNS03LX7	AL250036 Tetraodon
	20	127.4	12.7	729	28	BH284699	BH284699 CH230-49O
	21	127	12.7	599	28	BH486116	BH486116 BOGJZ09TF
	22	126.6	12.7	420	28	AZ782118	AZ782118 2M0022P01
	23	125.4	12.5	669	14	CA048466	CA048466 ssalsrkc0
	24	124.4	12.4	729	28	BZ041698	BZ041698 l1r23h12.
	25	123.4	12.3	633	28	BH604586	BH604586 BOHCF82TF
c	26	122.4	12.2	591	28	AZ633613	AZ633613 1M0488O23
c	27	122.4	12.2	1142	11	AK028099	AK028099 Mus muscu
	28	122	12.2	693	28	AZ650539	AZ650539 1M0520I14
	29	120.2	12.0	726	28	BZ015565	BZ015565 oeh87c12.
c	30	119.8	12.0	659	10	BB610246	BB610246 BB610246
	31	119.6	12.0	1100	9	AL528199	AL528199 AL528199
	32	118.6	11.9	736	29	BZ236616	BZ236616 CH230-251
c	33	118	11.8	680	28	AZ608625	AZ608625 1M0433A01
	34	118	11.8	712	14	CA047817	CA047817 ssalbrh01
	35	117.2	11.7	549	29	BZ285588	BZ285588 CH230-385
	36	117	11.7	740	28	BH328866	BH328866 CH230-93P
	37	115.8	11.6	564	28	BH515818	BH515818 BOHAC43TR
	38	115.6	11.6	711	12	BM950444	BM950444 UI-M-EH0p
	39	113.8	11.4	490	28	BH443125	BH443125 BOHRJ25TR
	40	113	11.3	711	29	BZ263814	BZ263814 CH230-420
	41	113	11.3	821	28	BH725237	BH725237 BOMHG42TF
	42	112.6	11.3	749	28	BH267859	BH267859 CH230-66L
c	43	111.6	11.2	660	28	BH474846	BH474846 BOHLI64TF
c	44	111.6	11.2	732	28	BZ060388	BZ060388 l1g98b12.
c	45	110.8	11.1	465	28	AZ800556	AZ800556 2M0058F10
	46	109.8	11.0	705	29	BZ470200	BZ470200 BONQQ85TF
	47	108.8	10.9	416	28	BZ179889	BZ179889 CH230-444
	48	108.8	10.9	438	9	AI676341	AI676341 etmEST008
	49	108.8	10.9	462	29	CNS04GRC	AL290001 Tetraodon
	50	108.8	10.9	528	28	BH700477	BH700477 BOMLB36TF
	51	108.4	10.8	794	28	BH544637	BH544637 BOGGJ62TF
	52	107.8	10.8	569	28	BH483727	BH483727 BOGCW50TF
	53	107.8	10.8	819	14	CB952830	CB952830 AGENCOURT
	54	106.6	10.7	677	28	BH677755	BH677755 BOHYJ15TR
	55	106.6	10.7	748	29	BZ497497	BZ497497 BONKB78TR
	56	106.4	10.6	457	9	AI676338	AI676338 etmEST008
c	57	104.4	10.4	566	14	CD347163	CD347163 EtESTef11
c	58	104.2	10.4	710	14	CB504454	CB504454 ssalplnb5
	59	102.6	10.3	4346	11	AK080786	AK080786 Mus muscu
c	60	102.4	10.2	551	29	BX173189	BX173189 Danio rer
c	61	102.4	10.2	787	29	BZ457666	BZ457666 BONRC33TF
	62	101.6	10.2	881	29	CNS02WKU	AL217191 Tetraodon
c	63	100.4	10.0	622	28	BH423105	BH423105 BOHJE88TF
c	64	100.2	10.0	592	29	BZ216441	BZ216441 CH230-323
c	65	100.2	10.0	702	29	BX242384	BX242384 Danio rer
	66	100	10.0	1201	13	BX426381	BX426381 BX426381
c	67	99.8	10.0	728	29	BX171642	BX171642 Danio rer
c	68	99.2	9.9	394	28	BH704622	BH704622 BOMBE06TR
	69	99.2	9.9	1094	29	CNS05B34	AL329305 Tetraodon
c	70	99	9.9	787	28	BH981983	BH981983 ode73b07.
c	71	98.6	9.9	734	14	CA056594	CA056594 ssalrgb53
c	72	98.4	9.8	310	28	BH115707	BH115707 RPCI-24-2
c	73	97.6	9.8	745	12	BI688458	BI688458 603316395
c	74	97.4	9.7	550	29	BX245766	BX245766 Danio rer
	75	97.2	9.7	393	10	BG561076	BG561076 EtESTed91
c	76	97.2	9.7	579	14	CD347167	CD347167 EtESTef11
c	77	96.8	9.7	477	28	AZ437267	AZ437267 1M0225L23

	78	96.6	9.7	404	14	CB811482	CB811482 AMGNNUC:M
c	79	96.2	9.6	490	29	DR42N9S	AL981355 Danio rer
	80	96.2	9.6	615	9	AI798117	AI798117 wh81d08.x
	81	95.8	9.6	1039	29	BZ224494	BZ224494 CH230-446
c	82	95.6	9.6	421	28	AZ857141	AZ857141 2M0161O23
c	83	95	9.5	839	29	BZ425593	BZ425593 BONIO62TF
c	84	95	9.5	2933	11	AK089803	AK089803 Mus muscu
	85	94.8	9.5	580	14	CA054386	CA054386 ssalrgb51
c	86	94.8	9.5	629	28	AZ364199	AZ364199 1M0110M16
	87	94.6	9.5	498	14	CB712482	CB712482 AMGNNUC:N
	88	94.6	9.5	656	28	AZ282406	AZ282406 RPCI-23-1
c	89	94.4	9.4	621	28	BZ180072	BZ180072 CH230-302
c	90	94.4	9.4	635	10	BB657463	BB657463 BB657463
c	91	94.4	9.4	1673	11	AK084134	AK084134 Mus muscu
	92	93.8	9.4	1152	29	CNS024NL	AL181002 Tetraodon
c	93	93.2	9.3	490	28	BH731712	BH731712 BOMAD94TF
c	94	93.2	9.3	737	14	CA063770	CA063770 ssalrgb53
	95	93.2	9.3	1057	29	CNS04CCM	AL284287 Tetraodon
c	96	92.8	9.3	461	29	FR0038988	AL126486 Fugu rubr
	97	92.6	9.3	1200	9	AL562271	AL562271 AL562271
	98	92.2	9.2	622	10	BB651343	BB651343 BB651343
	99	92.2	9.2	661	28	BH325278	BH325278 CH230-45J
c	100	92.2	9.2	760	28	AZ501999	AZ501999 1M0341N01
c	101	91.4	9.1	923	29	CNS03BUE	AL236975 Tetraodon
c	102	91.2	9.1	327	12	BM305533	BM305533 EtESTee32
	103	91.2	9.1	629	29	BZ452930	BZ452930 BONFM52TF
	104	91.2	9.1	766	29	CC324930	CC324930 TAM32-2J2
	105	90.8	9.1	485	28	BH729559	BH729559 BOMAZ10TR
c	106	90.6	9.1	705	14	CA352932	CA352932 624351 NC
c	107	90	9.0	1201	9	AL534303	AL534303 AL534303
	108	90	9.0	1201	13	BX405072	BX405072 BX405072
	109	89.2	8.9	602	28	AZ798563	AZ798563 2M0055M23
c	110	89.2	8.9	612	29	CNS02TXS	AL213769 Tetraodon
c	111	89.2	8.9	646	28	BH685737	BH685737 BOHXZ76TF
	112	89	8.9	474	9	AI226677	AI226677 uj13d03.y
	113	89	8.9	475	9	AI225652	AI225652 uj13h11.y
	114	89	8.9	526	9	AI786639	AI786639 uj17g02.y
	115	89	8.9	528	14	CA042084	CA042084 ssalplnb5
	116	89	8.9	569	9	AI786538	AI786538 uj17b08.y
	117	89	8.9	585	28	BH508316	BH508316 BOGUY28TR
	118	89	8.9	604	9	AI317436	AI317436 uj22f12.y
c	119	89	8.9	778	28	BH645114	BH645114 BOHVI77TR
c	120	89	8.9	833	29	BZ479906	BZ479906 BONIP76TR
c	121	88.8	8.9	438	9	AI676341	AI676341 etmEST008
c	122	88.8	8.9	550	28	AZ500983	AZ500983 1M0339E09
c	123	88.4	8.8	672	14	CB499018	CB499018 ssalkhc00
	124	88.2	8.8	1200	13	BX447644	BX447644 BX447644
c	125	87.8	8.8	711	12	BM950444	BM950444 UI-M-EH0p
c	126	87.8	8.8	799	14	CB499552	CB499552 ssalsrkc0
	127	87.6	8.8	270	28	BH562669	BH562669 BOHPE89TF
c	128	87.6	8.8	610	29	CNS04LLQ	AL296279 Tetraodon
	129	87.6	8.8	675	28	BH420110	BH420110 BOGJL74TF
c	130	87.4	8.7	833	29	CNS04472	AL273719 Tetraodon
	131	87.2	8.7	508	10	BF459386	BF459386 UI-M-BZ1-
	132	87.2	8.7	519	10	BF459383	BF459383 UI-M-BZ1-
c	133	87	8.7	673	14	CA052751	CA052751 ssalrgb51
c	134	86.6	8.7	599	28	BH652284	BH652284 BOHVU42TF
c	135	86.4	8.6	774	28	BZ143693	BZ143693 CH230-266
	136	86.4	8.6	961	29	CNS041D4	AL270049 Tetraodon
c	137	85.8	8.6	694	28	BH923991	BH923991 odh89f03.
c	138	85.8	8.6	1201	13	BX426381	BX426381 BX426381
	139	85.6	8.6	923	29	CNS03BUE	AL236975 Tetraodon
	140	85	8.5	261	28	AZ858629	AZ858629 2M0163G21
	141	83.8	8.4	212	28	BH248003	BH248003 BOGAC13TF
	142	83.6	8.4	721	28	BH484515	BH484515 BOHAX34TF
	143	83.4	8.3	911	29	CNS026IE	AL183407 Tetraodon
c	144	83.2	8.3	766	29	CC324930	CC324930 TAM32-2J2
c	145	83	8.3	354	28	BH515809	BH515809 BOHAC43TF
	146	83	8.3	511	10	BE028804	BE028804 EtESTea84
	147	83	8.3	1201	9	AL565450	AL565450 AL565450
	148	82.8	8.3	598	28	BH606983	BH606983 BOHBP37TF
	149	82.8	8.3	701	28	AZ363103	AZ363103 1M0108B02
	150	82.6	8.3	1860	11	AK082647	AK082647 Mus muscu
	151	82.4	8.2	496	10	BG561945	BG561945 EtESTee11
c	152	82.4	8.2	646	28	AZ962905	AZ962905 2M0231G22
c	153	82	8.2	393	10	BG561076	BG561076 EtESTed91
	154	82	8.2	872	29	CNS02859	AL185526 Tetraodon
c	155	81.6	8.2	463	29	BZ217730	BZ217730 CH230-323
c	156	81.6	8.2	849	13	BX462111	BX462111 BX462111
	157	81.6	8.2	1115	13	BX426372	BX426372 BX426372
	158	81.4	8.1	407	28	AZ847768	AZ847768 2M0148B03
c	159	81.2	8.1	753	28	BH531019	BH531019 BOGYM48TF
	160	81	8.1	327	12	BM305533	BM305533 EtESTee32
c	161	81	8.1	541	9	AI722296	AI722296 fc32b07.x
	162	80.6	8.1	855	28	BH599667	BH599667 BOGCE95TF
	163	80.4	8.0	650	29	BX214340	BX214340 Danio rer
c	164	80.4	8.0	713	28	BH547638	BH547638 BOHAC59TF
c	165	80.4	8.0	1055	29	CNS05EB2	AL333479 Tetraodon
	166	80.4	8.0	1201	13	BX405071	BX405071 BX405071
c	167	80.2	8.0	598	29	FR0029401	AL025772 Fugu rubr
c	168	80	8.0	457	9	AI676338	AI676338 etmEST008
	169	80	8.0	941	29	CNS046OU	AL276951 Tetraodon
	170	79.8	8.0	388	28	BH517322	BH517322 BOGXO85TR
c	171	79.8	8.0	1005	29	CNS05RJ5	AL350618 Tetraodon
c	172	79.8	8.0	1201	13	BX405071	BX405071 BX405071
c	173	79.4	7.9	719	28	AZ641572	AZ641572 1M0504K12
	174	79.4	7.9	743	29	BZ256408	BZ256408 CH230-334
	175	79.2	7.9	272	28	BH738439	BH738439 BOMAF71TR
	176	79.2	7.9	574	29	CNS0243A	AL180271 Tetraodon
	177	79.2	7.9	696	13	BU608109	BU608109 UI-CF-FN0
c	178	79	7.9	546	29	CNS03QQ0	AL256257 Tetraodon
	179	79	7.9	598	29	BZ521591	BZ521591 BOKAE54TF
	180	79	7.9	654	28	BH430125	BH430125 BOHEN70TF
c	181	78.8	7.9	644	29	BZ204163	BZ204163 CH230-351
c	182	78.6	7.9	1200	9	AL562271	AL562271 AL562271
	183	78.2	7.8	788	28	AZ643449	AZ643449 1M0506J21
	184	78	7.8	585	28	BH550521	BH550521 BOGKO58TR
c	185	77.8	7.8	581	14	CD342277	CD342277 EtESTee50
	186	77.8	7.8	592	28	BZ165534	BZ165534 CH230-463
c	187	77.8	7.8	827	28	BH250471	BH250471 BOGAK54TR
	188	77.6	7.8	744	28	BZ070693	BZ070693 1kk29f04.
c	189	77.6	7.8	888	28	BH713235	BH713235 BOMNI36TR
c	190	77.4	7.7	512	28	BH692899	BH692899 BOMFU72TR
c	191	77.4	7.7	1115	13	BX426372	BX426372 BX426372
c	192	77.2	7.7	619	28	AZ422325	AZ422325 1M0200P09
	193	77.2	7.7	635	28	BH699843	BH699843 BOMEH82TR
c	194	77.2	7.7	795	28	BH466067	BH466067 BOGVB34TR
c	195	76.8	7.7	856	29	CNS01X64	AL171301 Tetraodon
c	196	76.8	7.7	1274	29	CC209173	CC209173 CH261-44E
c	197	76.6	7.7	420	28	AZ782118	AZ782118 2M0022P01
	198	76.6	7.7	579	14	CD347167	CD347167 EtESTef11
	199	76.6	7.7	686	29	BZ511023	BZ511023 BOMRE33TR
c	200	76.6	7.7	771	28	BH701880	BH701880 BOMIF87TF
c	201	76.4	7.6	514	28	AZ251453	AZ251453 RPCI-23-2
c	202	76.4	7.6	545	28	AZ015249	AZ015249 RPCI-23-3
	203	76.2	7.6	500	13	BX425995	BX425995 BX425995
c	204	76.2	7.6	1201	13	BX405072	BX405072 BX405072
	205	76	7.6	1004	29	BZ443698	BZ443698 BONDV44TR
	206	75.8	7.6	581	14	CD342277	CD342277 EtESTee50
c	207	75.8	7.6	680	10	BG454776	BG454776 NF105F11L
	208	75.8	7.6	683	28	AZ421714	AZ421714 1M0200P05
	209	75.8	7.6	686	28	BH530855	BH530855 BOHEJ19TR
	210	75.8	7.6	850	13	BX404455	BX404455 BX404455
c	211	75.6	7.6	621	14	CB516309	CB516309 ssalrgb51
c	212	75.4	7.5	728	28	AZ458530	AZ458530 1M0262O07
	213	75	7.5	472	12	BI980396	BI980396 ft75h01.x
	214	74.8	7.5	510	28	AZ381359	AZ381359 1M0137G22
c	215	74.8	7.5	667	28	AZ707624	AZ707624 RPCI-23-2
c	216	74.8	7.5	685	29	CNS03YJF	AL266388 Tetraodon
c	217	74.8	7.5	725	28	AZ859279	AZ859279 2M0164N17
c	218	74.6	7.5	636	12	BM485474	BM485474 pgm1n.pk0
	219	74.6	7.5	789	29	BZ507343	BZ507343 BONFN66TR
	220	74.4	7.4	850	29	CNS03W36	AL263211 Tetraodon
c	221	74.4	7.4	1228	29	CC209174	CC209174 CH261-44E
c	222	74.2	7.4	1037	29	BZ230139	BZ230139 CH230-527
c	223	74	7.4	500	13	BX425995	BX425995 BX425995

	224	74	7.4	544	2	HSM095603 Bx501886 Homo sapi
	225	74	7.4	612	28	AZ501249 AZ501249 1M0340H03
c	226	74	7.4	755	29	BX158152 BX158152 Danio rer
c	227	73.8	7.4	580	29	DR24J5S AL735176 Danio rer
	228	73.6	7.4	261	14	CB498719 CB498719 ssalsrkc0
c	229	73.6	7.4	511	10	BE028804 BE028804 EtESTea84
c	230	73.6	7.4	671	28	AZ989593 AZ989593 2M0273D06
c	231	73.6	7.4	1066	29	CC206455 CC206455 CH261-138
	232	73.6	7.4	1201	13	BX356664 BX356664 BX356664
	233	73.4	7.3	521	28	AZ018805 AZ018805 RPCI-23-2
c	234	73.4	7.3	620	28	BH304233 BH304233 CH230-10C
	235	73.4	7.3	763	28	AZ705429 AZ705429 RPCI-23-2
	236	73.2	7.3	620	29	BZ240598 BZ240598 CH230-456
	237	73.2	7.3	676	29	BZ257915 BZ257915 CH230-479
	238	73.2	7.3	678	29	AG077026 AG077026 Pan trogl
c	239	73.2	7.3	692	28	BH577449 BH577449 BOGTF91TR
	240	73	7.3	350	28	BH716212 BH716212 BOHSX20TR
c	241	72.8	7.3	216	29	FR0026961 AL019790 F.rubripe
	242	72.8	7.3	358	10	BG562249 BG562249 EtESTee04
c	243	72.8	7.3	516	28	AZ494534 AZ494534 1M0329D21
c	244	72.8	7.3	693	28	BZ036855 BZ036855 oeh35h07.
c	245	72.6	7.3	1063	29	CNS04GZR AL290304 Tetraodon
	246	72.4	7.2	163	28	AZ764282 AZ764282 1M0560I07
	247	72.4	7.2	493	29	CNS07GXG AL610342 Anopheles
	248	72.4	7.2	1056	29	BZ494135 BZ494135 BONKT79TR
c	249	72.2	7.2	651	28	AZ314812 AZ314812 1M0031I18
	250	72.2	7.2	759	14	CA373131 CA373131 647164 NC
	251	72.2	7.2	860	29	SSC514813 AJ514813 Sus scrof
	252	72.2	7.2	866	29	CNS03NVU AL252579 Tetraodon
	253	72	7.2	444	14	CD344931 CD344931 EtESTee75
c	254	72	7.2	591	28	AZ611049 AZ611049 1M0436L01
	255	71.6	7.2	318	9	AW400672 AW400672 LamdiSest
	256	71.6	7.2	725	29	CNS01FOH AL142098 Anopheles
	257	71.6	7.2	992	29	CNS02NJX AL205494 Tetraodon
	258	71.4	7.1	431	10	BG724542 BG724542 EtESTed88
	259	71.4	7.1	450	10	BG724635 BG724635 EtESTed86
	260	71.4	7.1	487	10	BG516945 BG516945 EtESTed74
c	261	71.4	7.1	498	28	AZ633443 AZ633443 1M0488N09
c	262	71.4	7.1	620	29	CNS048OQ AL279539 Tetraodon
	263	71.4	7.1	650	28	BH494961 BH494961 BOGSS84TR
	264	71.4	7.1	651	28	AZ083794 AZ083794 RPCI-23-2
	265	71.4	7.1	977	29	CNS01VL4 AL169249 Tetraodon
	266	71.2	7.1	449	28	BZ103827 BZ103827 CH230-147
c	267	71.2	7.1	477	13	BX426373 BX426373 BX426373
c	268	71.2	7.1	576	12	BM582734 BM582734 170006872
	269	71	7.1	326	12	BM102557 BM102557 fu31g01.x
c	270	71	7.1	721	28	BH357652 BH357652 CH230-95B
c	271	71	7.1	1661	29	CC208142 CC208142 CH261-183
c	272	70.8	7.1	1201	29	CNS017M0 AL108210 Drosophil
	273	70.6	7.1	239	28	AZ586810 AZ586810 1M0394H05
	274	70.4	7.0	456	29	FR0030249 AL026618 Fugu rubr
	275	70.4	7.0	472	29	CC068214 CC068214 CSU-K33r.
	276	70.4	7.0	487	29	CC076917 CC076917 CSU-K33r.
	277	70.4	7.0	526	12	BI060010 BI060010 IL3-UT011
c	278	70.2	7.0	511	14	CD345061 CD345061 EtESTef64
c	279	70.2	7.0	644	28	BH588140 BH588140 BOHRL57TF
	280	70.2	7.0	660	29	DR48N15S AL984351 Danio rer
c	281	70.2	7.0	674	28	AZ990550 AZ990550 2M0274M15
	282	70	7.0	300	10	BG560916 BG560916 EtESTed95
	283	70	7.0	362	28	AZ516081 AZ516081 RPCI-11-8
	284	70	7.0	562	14	CD343982 CD343982 EtESTee90
c	285	70	7.0	663	28	AZ980406 AZ980406 2M0257L24
c	286	70	7.0	673	28	BH683564 BH683564 BOHXM85TR
c	287	70	7.0	776	28	BH592869 BH592869 BOGGW39TR
c	288	70	7.0	860	29	BZ211399 BZ211399 CH230-280
	289	69.8	7.0	570	28	AZ513198 AZ513198 1M0359O07
c	290	69.8	7.0	616	28	AZ406503 AZ406503 1M0175G13
	291	69.8	7.0	723	29	BX175762 BX175762 Danio rer
	292	69.8	7.0	852	28	BH719124 BH719124 BOHWT82TR
	293	69.8	7.0	1083	13	BQ941870 BQ941870 AGENCOURT
c	294	69.8	7.0	1200	13	BX447644 BX447644 BX447644
	295	69.6	7.0	601	29	CI12B2 AJ226232 Ciona int
	296	69.4	6.9	196	10	BG562268 BG562268 EtESTee04
c	297	69.4	6.9	481	29	BZ234457 BZ234457 CH230-427
c	298	69.4	6.9	520	14	CB126221 CB126221 K-EST0175
c	299	69.4	6.9	556	28	AZ847698 AZ847698 2M0148E23
	300	69.4	6.9	798	29	CNS025Z1 AL182710 Tetraodon

ALIGNMENTS

RESULT 1
AZ282406/c
LOCUS AZ282406 656 bp DNA linear GSS 27-JUL-2000
DEFINITION RPCI-23-138J12.TJ RPCI-23 Mus musculus genomic clone RPCI-23-138J12 , genomic survey sequence.
ACCESSION AZ282406
VERSION AZ282406.1 GI:9524192
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 656)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished
COMMENT Other_GSSs: RPCI-23-138J12.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 138 row: J column: 12
Seq primer: SP6
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .656
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-138J12"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). "
BASE COUNT 205 a 163 c 282 g 6 t
ORIGIN

Query Match 19.6%; Score 196.2; DB 28; Length 656;
Best Local Similarity 56.6%; Pred. No. 5.1e-35;
Matches 363; Conservative 0; Mismatches 278; Indels 0; Gaps 0;

```
Qy        105 GACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTG 164
              | | ||  |||||| ||   ||| |||||| ||       | | |   |||||| ||
Db        653 GCCTGTCTGCCTGCCTGTCTGCTTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCT 594

Qy        165 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTG 224
```

```
Db        593 GCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCT 534

Qy        225 GCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 284
              || |||  |   | ||||  ||  ||||||| ||| |||  ||  |||||  ||| |||
Db        533 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCT 474

Qy        285 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 344
              ||  |||||  ||| |||  ||  |||||  ||| |||  ||  |||||  ||| |||
Db        473 GCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCT 414

Qy        345 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 404
              |   |||||  ||| |||  ||  ||||||| ||| |||  ||  | |||| ||  |||
Db        413 GTCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCT 354

Qy        405 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCT 464
              ||  | |||| |||       | | ||  | |||| ||| |||       | | |    | ||
Db        353 GCCTGTCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGTCT 294

Qy        465 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 524
              |||||   || ||   ||  | |||   | ||||  |   | ||||  | ||||  ||
Db        293 GCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCT 234

Qy        525 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 584
              | ||||  | ||||  ||  | ||||  | ||||  |   | || |  | |  |  |
Db        233 GTCTGCCTGCCTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCAACATGCCT 174

Qy        585 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 644
              | |||   | ||||  |   | || |  | ||||  |   | ||||  | ||||  |
Db        173 GCCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 114

Qy        645 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTT 704
              | |||   | ||||  |   | ||||  | ||||  ||  | |||||||   |  ||| |
Db        113 GCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTGTCTGCCT 54

Qy        705 GCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTG 745
              || ||| |||  |   | | || || | | |  | ||  ||
Db         53 GCCTGCCTGCCTGCCTGCCTGCCTGACGGATTGTTGGTTTG 13
```

RESULT 2
AZ646178

```
LOCUS       AZ646178                759 bp    DNA     linear   GSS 14-DEC-2000
DEFINITION  1M0512A15F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0512A15 F, genomic survey sequence.
ACCESSION   AZ646178
VERSION     AZ646178.1  GI:11776574
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 759)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000  Std Error: 0.00
            Plate: 0512  row: A  column: 15
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 759.
```

```
FEATURES             Location/Qualifiers
     source          1. .759
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0512A15"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      126 a    167 c    177 g    288 t      1 others
ORIGIN
```

```
  Query Match             19.5%;  Score 195.2;  DB 28;  Length 759;
  Best Local Similarity   69.9%;  Pred. No. 9e-35;
  Matches  263;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy         72 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 131
              ||| ||| |||| ||| || ||||||||     |   |   || ||||||||| ||| ||
Db        380 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCGTGCTTGCTTGCTTGCG 439

Qy        132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191
              ||||||||     |   |   || ||||||||     |   |   || |||||||| |||
Db        440 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCGTGCTTGCTTGCT 499

Qy        192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251
              || |||||||      |  .|   |  |||| |||||||  |   |  |||| ||  ||
Db        500 TGCTTGCTTGCGTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCGTGCG 559

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
              ||||||||||| ||| || ||||||||||| ||| || ||||||||||| ||  ||
Db        560 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCGTGCT 619

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
              |||||||||||  ||  || |||||||||||  |||  | |||||||||||  ||| ||
Db        620 TGCTTGCTTGCTTGCGTGCTTGCTTGCTTGCTTACGTGCTTGCTTGCTTGCTTGCT 679

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTA 431
              ||||||||||| ||| || ||||||||||| ||| || ||||||||      |   |
Db        680 AGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCGTGCTTGCTAGCTTGCTTGCT 739

Qy        432 GGGCTGCTTGCTTGCT 447
              |  ||||||||||||
Db        740 TGCTTGCTTGCTTGCT 755
```

RESULT 3
BZ249369/c

```
LOCUS       BZ249369                821 bp    DNA     linear   GSS 12-OCT-2002
DEFINITION  CH230-409C7.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-409C7, genomic survey sequence.
ACCESSION   BZ249369
VERSION     BZ249369.1  GI:23910598
KEYWORDS    GSS.
```

```
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 821)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-409C7.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 409  row: C  column: 7
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .821
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-409C7"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      274 a    172 c    223 g    152 t
ORIGIN

  Query Match            16.6%;  Score 165.6;  DB 29;  Length 821;
  Best Local Similarity  64.2%;  Pred. No. 6.1e-28;
  Matches  249;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        109 GTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGG 168
              ||  ||||| ||||  ||  ||||||||||      | | ||  ||||| ||||
Db        402 GTCTGCCTGTTTGCCTGCCTGCCTGCTTGCCTGTTTGCCTGTTTGCCTGTTTGCCTGCTT 343

Qy        169 GACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTT 228
              |    | |       |  ||   ||     |||| |     |   |   |  |||  ||||
Db        342 GCGTCTTGCTTGCTTGCCTGCTTGCTTGCCTTGTTTGCCTGCTTGCCTGCTTGCCTGCTT 283

Qy        229 GCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 288
              ||| |   | ||||| ||| || ||| ||||||||||||| |||||||||||||  ||
Db        282 GCTTGCTTGTCTGCTTGCTCGCTTGCCTGCTTGCTCGCTTGCCTGCTTGCTTGCCTGCCT 223

Qy        289 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 348
              |||||||||| ||| ||  || |||||||||||| ||| |||||||||| |||| ||
Db        222 GCCTGCTTGCTCGCTTGCCTGCTTGCTTGCTTGCTTGCCTGCTTGCCTGCTTGCCT 163

Qy        349 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 408
              || ||| |||||||||||| |||||||||| |||  ||  || |||| ||| ||  |||
Db        162 GCTTGCCTGCTTGCTCGCTTGCCTGCTTGCCTGCCTGCCTGCTTGCTCGCTCGCCTGCTC 103

Qy        409 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCT 468
              ||||||||||     | | |   | |||||||| ||||      |       ||| ||| |
Db        102 GCCTGCTTGCTTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCTTGCTCACTCGCTTGCTT 43

Qy        469 GATGGCTTGATAGTAGGGCTGCTGGGCT 496
```

```
              | | |||   | | ||    ||   |||
Db         42 GCTCGCTCATTTGAAGCAGGGCATTGCT 15


RESULT 4
BZ515928/c
LOCUS       BZ515928                684 bp    DNA     linear   GSS 16-DEC-2002
DEFINITION  BOMRD40TR BO_2_3_KB Brassica oleracea genomic clone BOMRD40,
            genomic survey sequence.
ACCESSION   BZ515928
VERSION     BZ515928.1  GI:27045003
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 684)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .684
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOMRD40"
                     /clone_lib="BO_2_3_KB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      126 a    237 c    247 g     74 t
ORIGIN

  Query Match            14.8%;  Score 148.2;  DB 29;  Length 684;
  Best Local Similarity  53.6%;  Pred. No. 6e-24;
  Matches  354;  Conservative    0;  Mismatches  303;  Indels    4;  Gaps    2;

Qy         70 GCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGG 129
              || || |||  | |||||   | || ||||  |  |  |    || |||| | |||||
Db        670 GCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTAC 611

Qy        130 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCG 189
              | || ||||| |  |  |    || |||| | |        |||   ||  | | |||
Db        610 GCAGCGTGCTG-GTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTAC 552

Qy        190 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGG 249
              || |||| | |       |||   |  | | |||    ||   | | | |||||
Db        551 GCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCCCGGTCGCTAC 492

Qy        250 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 309
              | || |||  | |||||   | || |||  | |||||   | || |||  | |||||
Db        491 GCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTAC 432

Qy        310 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 369
              | || |||  | |||||   | || |||  | |||||   | || |||  | |||||
Db        431 GCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTAC 372

Qy        370 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 429
              | || |||  | |||||   | || |||  | |||||   | || |||   |  |  |
Db        371 GCAGCGTGCCCGGTCGCTATGCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTAC 312

Qy        430 TAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTG 489
                |  |||  | | |||| |  |   |     | || || | | | | ||    |  ||
```

```
Db        311 GCAGCATGCCCGGTCGCTACGCAGCATG---CTCAGCGTGCTCGGTCGCTACGCAGCGTG 255

Qy        490 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 549
              || ||  ||||    |  |||| ||  ||||    |  |||| ||  ||||    |  ||
Db        254 CTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTG 195

Qy        550 CTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTC 609
              || ||  ||||    |  | || ||  ||||    |  |||| ||  ||||    |  |
Db        194 CTCGGTCGCTACGCAGAGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTG 135

Qy        610 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 669
              || ||  ||||    |  |||| ||  ||||  .  |  |||| ||  ||||    |  ||
Db        134 CTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTG 75

Qy        670 CTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTG 729
              || ||  ||||    |  |||  | | |||     || |||| | | | |  || || ||
Db         74 CTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCGACGCAGCGTG 15

Qy        730 C 730
              |
Db         14 C 14
```

```
RESULT 5
AZ836738
LOCUS       AZ836738               545 bp    DNA     linear   GSS 20-FEB-2001
DEFINITION  2M0131P07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0131P07 R, genomic survey sequence.
ACCESSION   AZ836738
VERSION     AZ836738.1  GI:13006646
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 545)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0131  row: P  column: 07
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 545.
FEATURES             Location/Qualifiers
     source          1. .545
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0131P07"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       76 a    151 c    111 g    207 t
ORIGIN
```

```
  Query Match           14.6%;  Score 145.6;  DB 28;  Length 545;
  Best Local Similarity  78.1%;  Pred. No. 2.2e-23;
  Matches  175;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        224 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 283
              | ||||||  |   |  |||| ||  ||||||||||||||  ||| || |||||| ||||
Db          9 GCCTTGCTTGCTTGCTTGCTTGCCTGCCTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCT 68

Qy        284 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 343
              |||  ||||||  |||||||| ||| || ||||||||||||  ||| ||||||||||||||
Db         69 TGCTTGCCTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGCTTGCTTGCT 128

Qy        344 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403
              |||  |||||||||||||| ||  || ||||||||||||  ||| || |||||| ||||
Db        129 TGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGCT 188

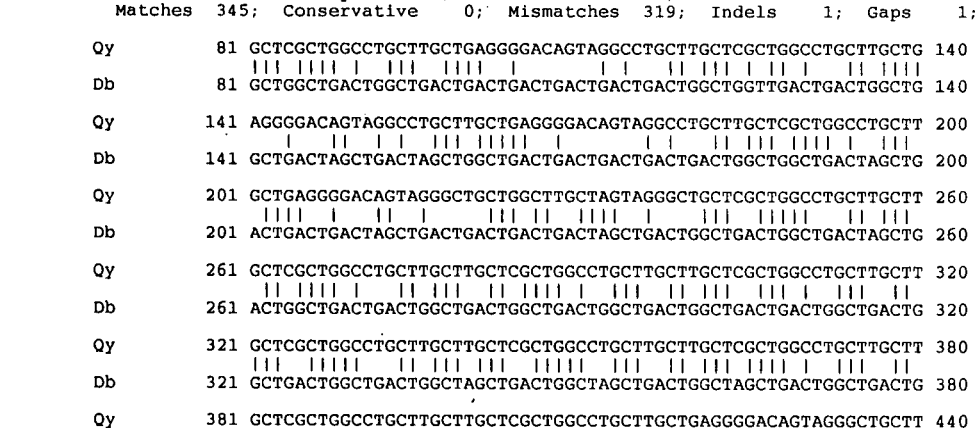
Qy        404 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCT 447
              ||| || ||| ||||     |   |   | ||||||||||||
Db        189 TGCTTGCTTGCCTGCTTGCTTGCTTGCTTGCCTGCTTGCTTGCT 232
```

```
RESULT 6
CB499552
LOCUS       CB499552               799 bp    mRNA    linear   EST 16-MAY-2003
DEFINITION  ssalsrkc017039 kidney Salmo salar cDNA, mRNA sequence.
ACCESSION   CB499552
VERSION     CB499552.1  GI:29310778
KEYWORDS    EST.
SOURCE      Salmo salar (Atlantic salmon)
  ORGANISM  Salmo salar
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
REFERENCE   1  (bases 1 to 799)
  AUTHORS   GRASP Consortium, Davidson,W.S., Koop,B.F. and
            http://web.uvic.ca/cbr/grasp.
  TITLE     A survey of Salmo salar transcripts from high complexity cDNA
            libraries
  JOURNAL   Unpublished
COMMENT     Contact: Koop BF
            Centre for Biomedical Research
            University of Victoria
            PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
            Tel: 250 472 4067
            Fax: 250 472 4075
            Email: bkoop@uvic.ca
            Centre for Biomedical Research, University of Victoria cDNA
            preparation and sequencing: Roberto Alberto, Marianne Beetz-Sargent
            , Maura Busby, Peter Hunt, Linda McKinnel, BF Koop. bioinformatics:
            Gordon D Brown.
FEATURES             Location/Qualifiers
     source          1. .799
                     /organism="Salmo salar"
                     /mol_type="mRNA"
                     /strain="McConnell"
                     /db_xref="taxon:8030"
                     /clone_lib="kidney"
                     /note="Vector: pBlueScriptIISK+; Library Creator:  Matthew
```

L Rise ; Atlantic salmon tissue contributors: Carlo Biagi, Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.), Simon Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery (Crofton, B.C.), Rachel Roper (University of Victoria)"

BASE COUNT 142 a 195 c 261 g 201 t
ORIGIN

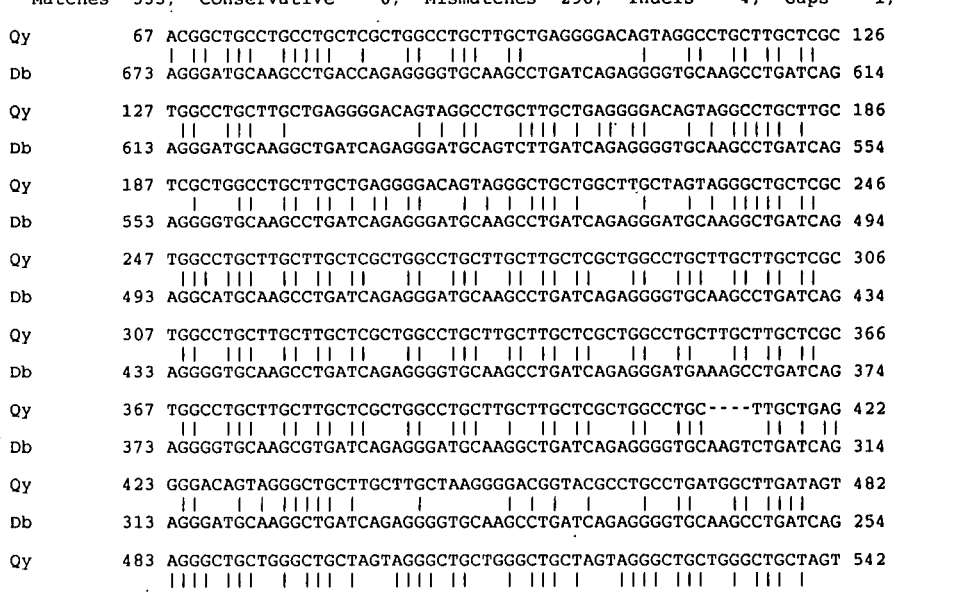
Query Match 14.3%; Score 142.6; DB 14; Length 799;
Best Local Similarity 51.9%; Pred. No. 1.2e-22;
Matches 345; Conservative 0; Mismatches 319; Indels 1; Gaps 1;



RESULT 7
BZ182690/c
LOCUS BZ182690 770 bp DNA linear GSS 11-OCT-2002
DEFINITION CH230-434H8.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone CH230-434H8, genomic survey sequence.
ACCESSION BZ182690
VERSION BZ182690.1 GI:23830009
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 770)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P. and Fraser,C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL Unpublished
COMMENT Other_GSSs: CH230-434H8.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230 (http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/or ering_information.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 434 row: H column: 8
Seq primer: SP6
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .770
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SsNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-434H8"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI; CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by Pieter de Jong"
BASE COUNT 120 a 296 c 141 g 213 t
ORIGIN

Query Match 14.2%; Score 142.2; DB 28; Length 770;
Best Local Similarity 52.6%; Pred. No. 1.5e-22;
Matches 335; Conservative 0; Mismatches 298; Indels 4; Gaps 1;



```
Db        253 AGGGGTGCAAGCCTGATCAGAGGGATGAAAGCCTGATCAGAGGGGTGCAAGCCTGATCAG 194

Qy        543 AGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGT 602
              |||| ||| ||||| |  |||| | |  ||||| |  |||| |||  | ||| |
Db        193 AGGGATGCAAGGCTGATCAGAGGGGTGCAAGGCTGATCAGAGGGATGCAAGCCTGATCAG 134

Qy        603 AGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGT 662
              |||| | |  | ||| |  |||| |||  | ||| |  |||| |||  | ||| |
Db        133 AGGGATGCAAGCCTGATCAGAGGGATGCAAGCCTGATCAGAGGGATGCAAGCCTGATCAG 74

Qy        663 AGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCT 699
              |||| |||  | ||| |  |||| |||  ||  | |
Db         73 AGGGATGCAAGCCTGATCAGAGGGATGCAAGCCTGGT 37


RESULT 8
BH276950
LOCUS       BH276950                 540 bp    DNA     linear   GSS 30-NOV-2001
DEFINITION  CH230-48M4.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
            CH230-48M4, genomic survey sequence.
ACCESSION   BH276950
VERSION     BH276950.1  GI:17189352
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 EcoRI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-48M4.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 48  row: M  column: 4
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-48M4"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 1"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      112 a    138 c    121 g    169 t
ORIGIN

  Query Match            14.1%;  Score 141.4;  DB 28;  Length 540;
  Best Local Similarity   81.9%;  Pred. No. 2.1e-22;
  Matches  163;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy        221 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 280
              ||| || |||| || ||  |||| | | || ||||||||||||| ||| || |||||||||

Db        329 GCTAGCCTGCTGGTGGGCTTGCTAGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 388

Qy        281 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 340
              ||| ||| || |||||||||||| ||| || |||||||||||| ||| || |||||||||
Db        389 GCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 448

Qy        341 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
              ||| ||| || |||||||||||| ||| || |||||||||||| ||| || |||||||||
Db        449 GCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTA 508

Qy        401 GCTCGCTGGCCTGCTTGCT 419
              ||| ||| || ||||| ||
Db        509 GCTTGCTTGCTTGCTTCCT 527


RESULT 9
BH484177
LOCUS       BH484177                 749 bp    DNA     linear   GSS 13-DEC-2001
DEFINITION  BOGHS47TR BOGH Brassica oleracea genomic clone BOGHS47, genomic
            survey sequence.
ACCESSION   BH484177
VERSION     BH484177.1  GI:17692281
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 749)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOGHS47TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .749
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGHS47"
                     /clone_lib="BOGH"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       96 a    221 c    254 g    178 t
ORIGIN

  Query Match            14.1%;  Score 140.6;  DB 28;  Length 749;
  Best Local Similarity   52.2%;  Pred. No. 3.5e-22;
  Matches  341;  Conservative    0;  Mismatches  304;  Indels    8;  Gaps    1;

Qy         73 GCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 132
              || |||  | |||||     || ||||  |  |  |  | || |||| | |||||
Db         96 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGAAGCGTGCTCGGTCGCTACGAA 155

Qy        133 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 192
              || ||||  |  |  |    || || | | |        ||  ||  | | |||  |
Db        156 GCGTGCTCGGTCGCTACGTAGCGTGATCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 215

Qy        193 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCT 252
              || |||| | |        ||   |  | | |||  |||  |  | |||||   |
Db        216 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTT 275

Qy        253 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 312
              || |||| | |||||    ||| |||| | |||||     || |||| | |||||
```

```
Db        276 GCGTGCTCGGTCGCTACTTTGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 335

Qy        313 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 372
              || |||| | |||||     || |||| | |||||     || |||| | |||||
Db        336 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 395

Qy        373 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAG 432
              || || | | |||||     || |||| | |||||     || ||||  |  |  |
Db        396 GCGTGGTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 455

Qy        433 GGCTGCTTGCTTGCTAAGGGGACGGTACG--------CCTGCCTGATGGCTTGATAGTAG 484
              |  |||| | | |||| |  |   |  ||          || || | | | | ||
Db        456 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 515

Qy        485 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 544
              |   ||| ||  ||||    |  |||| ||  ||||    |  |||| ||  ||||
Db        516 GCGAGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGGAGCGTGCTCGGTCGCTACGTT 575

Qy        545 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 604
                |||| ||  ||||    |  | || ||  ||||    |  |||| ||  ||||
Db        576 ACATGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 635

Qy        605 GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 664
              |  | ||  ||  ||||    |  |||| ||  ||||       |||| ||  ||||
Db        636 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAACGTGCTCGGTCGCTACGTA 695

Qy        665 GGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAG 717
              |  |||| ||  ||||    |  |||  | | |||   | || |||| | | |
Db        696 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCG 748


RESULT 10
BZ519739
LOCUS       BZ519739                712 bp    DNA     linear   GSS 16-DEC-2002
DEFINITION  BOMRL03TR BO_2_3_KB Brassica oleracea genomic clone BOMRL03,
            genomic survey sequence.
ACCESSION   BZ519739
VERSION     BZ519739.1  GI:27051945
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 712)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .712
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOMRL03"
                     /clone_lib="BO_2_3_KB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       83 a    238 c    243 g    148 t
ORIGIN

  Query Match          14.0%;  Score 140;  DB 29;  Length 712;
  Best Local Similarity   52.2%;  Pred. No. 4.7e-22;
  Matches  363;  Conservative    0;  Mismatches  325;  Indels    8;  Gaps    2;

Qy         73 GCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 132
              || |||  | |||||     || ||||  |  | ||    || |||| | ||||
Db         21 GCATGCTCGGTCGCTACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 80

Qy        133 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 192
              || |||| |  | ||    || |||| |    |  | ||    || |||| | ||||
Db         81 GCGTGCTCGGTCGCCACGTAGCGTGCTCG----GTCGCCACGTAGCGTGCTCGGTCGCCA 136

Qy        193 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCT 252
                 || ||||  |  | ||    |  |||| | | || |    |  |||||| | |||
Db        137 CGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCA 196

Qy        253 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 312
               | || |||||| | |||   | || ||||||| | |||   | || |||||| | |||
Db        197 CGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCA 256

Qy        313 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 372
               | || |||||| | |||   | || ||||||| | |||   | || |||||| | |||
Db        257 CGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCA 316

Qy        373 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC----TGCTTGCTGAGGGGACA 428
               | || |||||| | |||   | || |||||| | |||     || |||| |  | ||
Db        317 CGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCA 376

Qy        429 GTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCT 488
               |  |||| | | || | |  |   |  ||   |||  |||| |||| |   | |
Db        377 CGTAGCGTGCTCGGTCGCCACGTAGCGTGATCGGTCGCCACGTGGCGTGATCGGTCGCCA 436

Qy        489 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 548
               | |  || | |   | |   |||  |||| |   | |   | |  || | |   | |
Db        437 CGTAGCGTGATCGGTCGCCACGTGGCGTGCTCGGTCGCCACGTAGCGTGTTCGGTCGCCA 496

Qy        549 GCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 608
               | |  |||| |   | | | | |  |||| |   | |   | |  |||| |   | |
Db        497 CGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCA 556

Qy        609 CCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 668
               | ,|  |||| |   | |   | |  |||| |   | |   | |  |||| |   | |
Db        557 CGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCA 616

Qy        669 GCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTT 728
               | |  |||| |   | |     ||  ||| | | ||    | ||  |   || ||||
Db        617 CGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCTA 676

Qy        729 GCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACA 764
                 || ||   |  | || |  |   | |  |||
Db        677 CGTAGCGACCGTGCTTTCTTTAAAATCGATACGACA 712


RESULT 11
BH541157/c
LOCUS       BH541157                845 bp    DNA     linear   GSS 14-DEC-2001
DEFINITION  BOGLR58TR BOGL Brassica oleracea genomic clone BOGLR58, genomic
            survey sequence.
ACCESSION   BH541157
VERSION     BH541157.1  GI:17789521
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 845)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOGLR58TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
```

```
           Email: cdtown@tigr.org
           DNA is from a doubled haploid provided by Tom Osborn.
           Seq primer: TR
           Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .845
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGLR58"
                     /clone_lib="BOGL"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      200 a    307 c    254 g     84 t
ORIGIN

  Query Match             13.6%;  Score 136.2;  DB 28;  Length 845;
  Best Local Similarity   51.4%;  Pred. No. 3.7e-21;
  Matches  345;  Conservative    0;  Mismatches  318;  Indels    8;  Gaps    1;

Qy         73 GCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 132
              || |||  | |||||     || ||||  |  |  |    || |||| | |||||
Db        843 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 784

Qy        133 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 192
              || ||||  |  |  |    || |||| | |         ||   ||  | | |||   |
Db        783 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 724

Qy        193 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCT 252
              || |||| | |        ||   |   | | |||   |||   |   | |||||
Db        723 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 664

Qy        253 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 312
              || |||| | |||||     || |||| | |||||     || |||| | |||||
Db        663 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 604

Qy        313 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 372
              || |||| | |||||     || |||| | |||||     || |||| | |||||
Db        603 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 544

Qy        373 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAG 432
              || |||| | |||||     || |||| | |||||     || ||||  |  |  |
Db        543 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 484

Qy        433 GGCTGCTTGCTTGCTAAGGGGACGGTACGCCT--------GCCTGATGGCTTGATAGTAG 484
              |  |||| | |||| |  |   |  ||            || || | | | | ||
Db        483 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 424

Qy        485 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 544
              |  |||| ||  ||||     |  |||| ||  ||||     |  |||| ||  ||||
Db        423 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 364

Qy        545 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 604
              |  |||| ||  ||||     |  | || ||  || |     |  |||| ||  || |
Db        363 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCGACGTAGCGTGCTCGGTCGCGACGTA 304

Qy        605 GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 664
              |  | || ||  || |     |  | || ||  || |     |  |||| ||  || |
Db        303 GCGTGCTCGGTCGCGACGTAGCGTGCTCGGTCGCGACGTAGCGTGCTCGGTCGCGACGTA 244

Qy        665 GGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCC 724
              |  |||| ||  || |     |  ||| | | ||     | || |||| | | | |  |
Db        243 GCGTGCTCGGTCGCGACGTAGCGTGCTCGGTCGCGACGTAGCGTGCTCGGTCGCGACGTA 184

Qy        725 GCTTGCCTGCT 735
              || |||  | |
Db        183 GCGTGCTCGGT 173


RESULT 12
BH521799
LOCUS       BH521799               697 bp    DNA     linear   GSS 13-DEC-2001
DEFINITION  BOGXL17TF BOGX Brassica oleracea genomic clone BOGXL17, genomic
            survey sequence.
ACCESSION   BH521799
VERSION     BH521799.1  GI:17729884
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 697)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOGXL17TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .697
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGXL17"
                     /clone_lib="BOGX"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       84 a    232 c    231 g    150 t
ORIGIN

  Query Match             13.5%;  Score 135.4;  DB 28;  Length 697;
  Best Local Similarity   52.0%;  Pred. No. 5.3e-21;
  Matches  330;  Conservative    0;  Mismatches  301;  Indels    4;  Gaps    1;

Qy        117 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 176
              || |||||| | |||   | ||     |   |     || ||||  |  | ||
Db          3 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 62

Qy        177 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 236
              || |||| | ||||      || ||||  |  | ||    |  |||| | | || |
Db         63 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 122

Qy        237 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 296
              |  |||||| | |||   | || ||||||  | |||   | || ||||||  | |||   |
Db        123 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 182

Qy        297 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 356
              || |||||| | |||   | || ||||||  | |||   | || ||||||  | |||   |
Db        183 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 242

Qy        357 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC----T 412
              || |||||| | |||   | || ||||||  | |||   | || ||||||  | |||
Db        243 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 302

Qy        413 GCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATG 472
              || ||||  |  | ||    |  |||| | | || | |  |   |  ||   |||   ||
Db        303 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGATCGGTCGCCACGTG 362

Qy        473 GCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 532
              || |||| |   | |   | |  || | |   | |   |||  |||| |   | |   |
Db        363 GCGTGATCGGTCGCCACGTAGCGTGATCGGTCGCCACGTGGCGTGCTCGGTCGCCACGTA 422

Qy        533 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTG 592
              |  || | |   | |  |||| |   | | | | |  |||| |   | |   |
Db        423 GCGTGTTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 482
```

```
Qy        593 GGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 652
              |  ||||  |   | | | |  ||||  |   | |   | |  ||||  |   | |   |
Db        483 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 542

Qy        653 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTT 712
              |  ||||  |   | |   | |  ||||  |   | |     ||  ||| | | ||    |
Db        543 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 602

Qy        713 GCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCT 747
              ||  |   || ||||    || ||   |  | ||
Db        603 GCGTGCTCGGTCGCTACGTAGCGACCGTGCTTTCT 637


RESULT 13
BZ221738
LOCUS       BZ221738                765 bp    DNA     linear   GSS 11-OCT-2002
DEFINITION  CH230-372D22.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-372D22, genomic survey sequence.
ACCESSION   BZ221738
VERSION     BZ221738.1  GI:23880096
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 765)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-372D22.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 372  row: D  column: 22
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .765
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-372D22"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      213 a    137 c    302 g    113 t
ORIGIN

  Query Match             13.4%;  Score 133.6;  DB 29;  Length 765;
  Best Local Similarity   51.6%;  Pred. No. 1.4e-20;
  Matches  358;  Conservative    0;  Mismatches  329;  Indels    7;  Gaps    2;

Qy          5 GGGAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGG 64
              ||| | | ||  |   || |||  || ||     |    | ||| |  ||     |
Db         74 GGGTACAAGCCTGATCAGAGGGATGCAAGGCTGATCAGAGGGATGCAAGCCTGATCAGAG 133

Qy         65 GTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 124
              |     |||||       |   ||  |  || | || ||   | | |||  | |
Db        134 GGGTACAAGCCTGATCAGAGGGGTGCAAGGCTGATCAGAGGGGTGCAAGCCAGATCAGAG 193

Qy        125 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 184
              |  ||  |  || | || ||   | | |||||   | | || ||   | | | ||| |
Db        194 GGGTGCAAGGCTGATCAGAGGGATGCAAGCCTG---GATCAGAGGGGTGCAAGGCTGATC 250

Qy        185 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTC 244
              |   ||  || || | || ||   | | | ||| |     |    | | ||| ||
Db        251 AGAGGGATGCAAGCCTGATCAGAGGGGTGCAAGCCTGATCAGAGGGGTACAAGCCTGATC 310

Qy        245 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 304
              ||  |||  |  || ||   ||  | |  || || ||   ||  ||| |  || ||
Db        311 AGAGGGATGCAAGGCTGATCAGAGGGGTACAAGCCTGATCAGAGGGGTGCAAGGCTGATC 370

Qy        305 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 364
              ||  |||  || || ||   ||  |||  || || ||   ||  |||  || || ||
Db        371 AGAGGGATGCAAGCCTGATCAGAGGGATGCAAGCCTGATCAGAGGGATGCAAGCCTGATC 430

Qy        365 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC----TTGCTG 420
              ||  |||  || || ||   ||  | |  || || ||   ||  |||      || |
Db        431 AGAGGGATGCAAGCCTGATCAGAGGGGTACAAGCCTGATCAGAGGGATGCAAGGCTGATC 490

Qy        421 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 480
              || ||   | | | |||  |      |      | | |  |    |  ||   || ||||
Db        491 AGAGGGGTGCAAGCCTGATCAGAGGGGTGCAAGGCTGATCAGAGGGATGCAAGCCTGATC 550

Qy        481 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 540
              ||||  |||  |||||  |   |||| |||  |||||  |   |||| |||  | ||| |
Db        551 AGAGGGGTGCAAGGCTGATCAGAGGGATGCAAGGCTGATCAGAGGGGTGCAAGCCTGATC 610

Qy        541 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 600
              ||||  |||  |||||  |   |||| | |  | |||  |   |||| |||  |||||  |
Db        611 AGAGGGGTGCAAGGCTGATCAGAGGGGTGCAAGCCTGATCAGAGGGGTGCAAGGCTGATC 670

Qy        601 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 660
              ||||  | |  | |||  |   ||||  |||  |||||  |   | || |||  |||||  |
Db        671 AGAGGGGTGCAAGCCTGATCAGAGGGGTGCAAGGCTGATCAGAAGGGTGCAAGGCTGATC 730

Qy        661 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGC 694
              ||||  |||  |||||  |   ||||  |||  ||
Db        731 AGAGGGGTGCAAGGCTGATCAAAGGGGTGCAAGC 764


RESULT 14
AZ828483/c
LOCUS       AZ828483                481 bp    DNA     linear   GSS 20-FEB-2001
DEFINITION  2M0105A06R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0105A06 R, genomic survey sequence.
ACCESSION   AZ828483
VERSION     AZ828483.1  GI:12998391
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 481)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
```

```
           Insert Length: 10000   Std Error: 0.00
           Plate: 0105  row: A  column: 06
           Seq primer: CACACAGGAAACAGCTATGACC
           Class: plasmid ends
           High quality sequence stop: 481.
FEATURES             Location/Qualifiers
     source          1. .481
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0105A06"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      148 a    137 c    116 g     80 t
ORIGIN

  Query Match             13.2%;  Score 132.4;  DB 28;  Length 481;
  Best Local Similarity   76.2%;  Pred. No. 2.4e-20;
  Matches  163;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy        202 CTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTG 261
              |||| |  |||||  || |    ||||||| |   |   ||| |||||| |||| |||||
Db        476 CTGAAGCTACAGTCAGGGTCAGAGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTG 417

Qy        262 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 321
              || ||| ||  |||||||| ||| |||||| |||| |||||||| ||| ||  |||||||
Db        416 CTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCTGG 357

Qy        322 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 381
              || |||||| |||| ||||||| ||| ||  ||||||| ||| |||||| |||| |||||
Db        356 CTTGCTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTG 297

Qy        382 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 415
              || ||| ||  ||||||| ||| |||||| ||||
Db        296 CTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCT 263


RESULT 15
BH537125
LOCUS       BH537125                597 bp    DNA     linear   GSS 14-DEC-2001
DEFINITION  BOHCS81TR BOHC Brassica oleracea genomic clone BOHCS81, genomic
            survey sequence.
ACCESSION   BH537125
VERSION     BH537125.1  GI:17772878
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 597)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOHCS81TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .597
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHCS81"
                     /clone_lib="BOHC"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       34 a    180 c    237 g    146 t
ORIGIN

  Query Match             13.2%;  Score 132;  DB 28;  Length 597;
  Best Local Similarity   53.1%;  Pred. No. 3.1e-20;
  Matches  305;  Conservative    0;  Mismatches  265;  Indels    4;  Gaps    1;

Qy         73 GCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 132
              || |||  | ||||||    || ||||  |  |       || |||| | ||||||
Db          5 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 64

Qy        133 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 192
              || ||||  |  |       || |||| | |    |   ||  ||  | | |||   |
Db         65 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 124

Qy        193 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCT 252
              || |||| | |    |   ||   |  | | |||   |||   |   | ||||||
Db        125 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 184

Qy        253 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 312
              || |||| | ||||||    || |||| | ||||||    || |||| | ||||||
Db        185 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 244

Qy        313 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 372
              || |||| | ||||||    || |||| | ||||||    || |||| | ||||||
Db        245 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 304

Qy        373 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAG 432
              || |||| | ||||||    || |||| | ||||||    || ||||  |  |
Db        305 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 364

Qy        433 GGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTG 492
              |  |||| | | |||  |  |   |  ||   ||    | || || |    ||  ||||
Db        365 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCT----CGGTCGCTG 420

Qy        493 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 552
              |  ||  |   ||  |||| |  ||  |   ||  |||| |  ||  |   ||  |||
Db        421 CGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTG 480

Qy        553 GGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTG 612
              |  ||  |   ||  ||| |  ||  |   ||  |||| |  ||  |   ||  |||
Db        481 CGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTG 540

Qy        613 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 646
              |  ||  |   ||  ||| |  ||  ||||||
Db        541 CGTAGCGTGCTCGGTCGCTACGTAGCGCGTAGGG 574


RESULT 16
BX333427/c
LOCUS       BX333427                1053 bp    mRNA    linear   EST 01-MAY-2003
```

```
DEFINITION  BX333427 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DC025YL15 5-PRIME, mRNA sequence.
ACCESSION   BX333427
VERSION     BX333427.1  GI:30312213
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1053)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 5205.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DC025CF08QP1&cluster=5205.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DC025CF08QP1.
FEATURES             Location/Qualifiers
     source          1. .1053
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DC025YL15"
                     /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      255 a    462 c    300 g     36 t
ORIGIN

  Query Match             13.2%;  Score 132;  DB 13;  Length 1053;
  Best Local Similarity   59.9%;  Pred. No. 3.6e-20;
  Matches  454;  Conservative    0;  Mismatches  280;  Indels   24;  Gaps   13;

Qy          69 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 128
               ||||| ||||  ||| ||||||  ||| |||| |  | | |  |||  ||| ||| ||||
Db         907 GGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTG-GCTGGCTGCTGGCTGGCTGGCTGGCTG 849

Qy         129 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 188
               ||  ||| ||||  ||  | |   | |||  ||    |  | | |  |||  ||| |||
Db         848 GCTGGCTGGCTGCTGG--CTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTG 791

Qy         189 GCTGGCCTGCTTGCTGAGGG------GACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGC 242
               ||||||  ||| ||||  ||      | | |   |||||   |  |||| |  ||   ||
Db         790 GCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGC 731

Qy         243 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 302
               | ||||||  ||| || |||| ||||||  ||| ||| ||| |||||  |||||  ||| ||
Db         730 TGGCTGGCTGGCTGGC-TGCTGGCTGGCTGGCTGGCTGGCTGG-CTGCTGGCTGGC 673

Qy         303 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 362
               | ||||||  ||| ||| ||| ||||||  ||| ||| ||| |||||  |||  |||| ||
Db         672 TGGCTGGCTGGCTGGCTGGCT-GCTGGCTGGCTGGCTGGCTGGCTGG-CTGGCTGCTGGC 615

Qy         363 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAG 422
               | ||||||  ||| ||| ||| ||| ||  ||| ||| ||| ||||||  ||| ||||
Db         614 TGGCTGGCTGGCTGGCTGGCTGGCT-GCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCT 556

Qy         423 GG--GACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 480
               ||  | | |   |||||  ||  ||      ||  ||   ||  ||  | |||||  | |
Db         555 GGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 496

Qy         481 GTAG--GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 538
               |||| |  |||||   |||||   |   |||||||||   ||| |  ||   |||||   ||
Db         495 CTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGC 436

Qy         539 TAGTAG--GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCT 596
               |  | |  |||||   |||||   |   |||| ||||   ||| |  ||   |||||
Db         435 TGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGG 376

Qy         597 GCTAGTAG--GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 654
               |||  | |  ||||    |||||   |   |||||||||   ||| |  ||   |||||
Db         375 GCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGC 316

Qy         655 CTGCTAGTAG--GGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTT 712
                 |||  | |  |||||   |||||   |   ||||||  |||||||  ||| ||| |||
Db         315 TGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTG 256

Qy         713 GCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGGG 772
               |||  |  ||  | |   |  |||        |  ||  |||  |   |     ||   |  |
Db         255 GCTGGCTGGCCTGGGCGGATGCACTGGCAGTGGCAATGCTGGCCGTCTGGCAACTTTTCG 196

Qy         773 CTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAG 810
                  ||   | | |||    || |||   | |  || ||
Db         195 TCTCCCCGCGGATGCCCTGGACGGCCCTCCACTGCGAG 158


RESULT 17
CNS03LWH/c
LOCUS       CNS03LWH                 894 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            036N09 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL250010
VERSION     AL250010.1  GI:7971022
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 894)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .894
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="036N09"
```

```
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG036CG05SP1~end :
                     PUC-Ori"
BASE COUNT      243 a    396 c    221 g     32 t      2 others
ORIGIN

  Query Match             13.1%;  Score 130.8;  DB 29;  Length 894;
  Best Local Similarity   60.0%;  Pred. No. 6.6e-20;
  Matches  468;  Conservative    0;  Mismatches  287;  Indels   25;  Gaps   14;

Qy          69 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 128
               ||||| |||| ||| |||||| ||| |||| | | | | ||| ||| ||| ||||
Db         872 GGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTG-GCTGGCTGCTGGCTGGCTGGCTGGCTG 814

Qy         129 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 188
               || ||| |||| || | |  | ||| ||    | | |.| ||| ||| |||
Db         813 GCTGGCTGGCTGCTGG--CTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTG 756

Qy         189 GCTGGCCTGCTTGCTGAGGG------GACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGC 242
               |||||| ||| |||| ||      | | |   |||||   |  |||| |  ||   ||
Db         755 GCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGC 696

Qy         243 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 302
               | |||||| ||| || |||| |||||| ||| ||| ||| ||||| ||||| ||| ||
Db         695 TGGCTGGCTGGCTGGC-TGCTGGCTGGCTGGCTGGCTGGCTGGCTGG-CTGCTGGCTGGC 638

Qy         303 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 362
               | |||||| ||| ||| ||| |||||| ||| ||| ||| ||||| ||| |||| ||
Db         637 TGGCTGGCTGGCTGGCTGGCT-GCTGGCTGGCTGGCTGGCTGGCTGG-CTGGCTGCTGGC 580

Qy         363 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAG 422
               | |||||| ||| ||| ||| ||  ||| ||| ||| ||||||  ||| ||||
Db         579 TGGCTGGCTGGCTGGCTGGCTGGCT-GCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCT 521

Qy         423 GG--GACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 480
               ||  | | |  ||||| ||  ||      ||  ||   ||  ||  | ||||| | |
Db         520 GGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTG 461

Qy         481 GTAG--GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 538
               | |  |||||   |||||   |  ||||||||||   ||| |  ||  |||||   ||
Db         460 CTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGC 401

Qy         539 TAGTAG--GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCT 596
               |  | |  |||||   |||||   |   |||| ||||   ||| |  ||   |||||
Db         400 TGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGCTG 341

Qy         597 GCTAGTAG--GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 654
               |||  | |  ||||    |||||   |  |||||||||   ||| |  ||   |||||
Db         340 GCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGC 281

Qy         655 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGC 714
                 |||  | |||||   |||||   |  ||||| |||||||||| ||| ||| ||| ||
Db         280 TGGCTGCT--GGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGC 223

Qy         715 TAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGGGCT 774
               | |   ||| ||| ||  |||    ||   ||||    ||||   |       |  ||||
Db         222 T-GGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCT 164

Qy         775 GCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACCCGGGCCTC 834
               | | | | |     ||   ||  | | |   | | | |   |     ||     || ||||
Db         163 GGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGATGGACCTC 104


RESULT 18
DR37P5T/c
LOCUS       DR37P5T                  640 bp    DNA     linear   GSS 22-NOV-2002
DEFINITION  Danio rerio genomic clone DKEY-37P5, genomic survey sequence.
ACCESSION   AL975317
VERSION     AL975317.1  GI:25181115
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 640)
  AUTHORS   Humphray,S.J., Huckle,E. and Hunt,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2002) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 37P5. 37P5 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene.
            Further details: http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .640
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-37P5"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
BASE COUNT      299 a    232 c     55 g     54 t
ORIGIN

  Query Match             13.0%;  Score 129.8;  DB 29;  Length 640;
  Best Local Similarity   55.4%;  Pred. No. 1e-19;
  Matches  292;  Conservative    0;  Mismatches  232;  Indels    3;  Gaps    2;

Qy         220 TGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 279
               ||||  | || | ||     || | | | || || ||  ||||| | |  | || ||| |
Db         638 TGCTTACCTGGTTGTTTCCTTGGTTGTTAGCTTGGTTAGTTGCTTGATTACTTGGTTGGT 579

Qy         280 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 339
               || | | | |  || ||||||    |||    || ||  ||||| ||| || |||||| |
Db         578 TGTTTGATTGGTTGGTTGCTTTGATGCTTAGTTGGTTAGTTGCTTGCTTGCTTGCTTGAT 519

Qy         340 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 399
               | |  |||| |  ||| ||  | ||| || |||||||||| | | || |  | | | |
Db         518 AGTTGACTGG-TTAGTTGGTTAGTTGCTTGCTTGCTTGCTTGATAGTTGACTGGTTAGTT 460

Qy         400 TGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAG--GGCTGCTTGCTTGCTAAGGGGACGG 457
               | | |||||   | | | ||    |  || |  || |||||||||| || |  |
Db         459 GGTTGGCTGGTTGGTTGGTTGGTTAGTTGGTTGTTGGTTGCTTGCTTGGTATGTTGGTTA 400

Qy         458 TACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTG 517
               || ||  || | ||| | | |   || || | || || | |   || || |  | |
Db         399 GGTGCTTGGTTGGTTGCTAGCTTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTAGTTA 340

Qy         518 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTG 577
               || |   || || | |  || |     || ||    | || | ||  || |  | || ||
Db         339 CTTGGTTGGTTGGTTGATTGGTTTCTTGGTTGGATAGTTGGTTGTTTGGTTGGTTGGTTG 280

Qy         578 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTG 637
               |     |  || | || ||||  |   || | ||    || | |   || || |  | ||
Db         279 GTTAGTTGTTTGGTTGGTTGCTTGGTAGGTTGCTTACTTGGTTGGTTGGATGGTTGGTTG 220

Qy         638 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGG 697
               | |   || || | || || | |   || |  | || |||| |  ||| ||  || | |
Db         219 GTTGGTTGGTTGGTTGGTTGGTTGGTTTGTTGGTTGCTTGGTGGGTTGAGTGGTTG 160

Qy         698 CTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCT 744
               |||||| |||| ||| |     ||  | | |  || |     ||||
Db         159 GTTGCTTACTTGGTTGGTTTCTTGGTTGGTCGGTTGGTTGGTTGGCT 113


RESULT 19
CNS03LX7
LOCUS       CNS03LX7                 743 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            036P11 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL250036
VERSION     AL250036.1  GI:7971048
KEYWORDS    GSS; genome survey sequence.
```

```
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 743)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .743
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="036P11"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG036CH06SP1-end :
                     PUC-Ori"
BASE COUNT        0 a    191 c    360 g    192 t
ORIGIN

  Query Match           12.9%;  Score 129.4;  DB 29;  Length 743;
  Best Local Similarity   61.5%;  Pred. No. 1.3e-19;
  Matches  448;  Conservative    0;  Mismatches  256;  Indels   25;  Gaps   14;

Qy         69 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 128
              ||||| ||||  ||| ||||||  ||| |||| |  | | |  |||  ||| ||| ||||
Db         22 GGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTG-GCTGGCTGCTGGCTGGCTGGCTGGCTG 80

Qy        129 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 188
              ||  ||| ||||  ||  | |   | |||  ||    |  | | |  |||  ||| |||
Db         81 GCTGGCTGGCTGCTGG--CTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTG 138

Qy        189 GCTGGCCTGCTTGCTGAGGG------GACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGC 242
              ||||||  ||| ||||  ||      | | |   |||||   |  |||| |  ||   ||
Db        139 GCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGC 198

Qy        243 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 302
              | ||||||  ||| || |||| ||||||  ||| ||| ||| ||||| ||||| ||| ||
Db        199 TGGCTGGCTGGCTGGC-TGCTGGCTGGCTGGCTGGCTGGCTGGTGG-CTGCTGGCTGGC 256

Qy        303 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 362
              | ||||||  ||| ||| ||| ||||||  ||| ||| ||| ||||| |||  |||| ||
Db        257 TGGCTGGCTGGCTGGCTGGCT-GCTGGCTGGCTGGCTGGCTGGCTGG-CTGGCTGCTGGC 314

Qy        363 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAG 422
              | ||||||  ||| ||| ||| ||| ||  ||| ||| ||| ||||||  ||| ||||
Db        315 TGGCTGGCTGGCTGGCTGGCTGGCT-GCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 373

Qy        423 GG--GACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 480
              ||  | | |   |||||  ||  ||       ||  ||   ||  ||  | ||||| | |
Db        374 GGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTG 433

Qy        481 GTAG--GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 538
                | |  |||||   |||||   |  |||||||||   ||| |  ||   |||||  ||
Db        434 CTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGC 493

Qy        539 TAGTAG--GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCT 596
              | | |  |||||   |||||    |   |||| ||||   ||| |  ||   |||||
Db        494 TGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTG 553

Qy        597 GCTAGTAG--GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 654
              |||  | |  ||||    |||||    |   |||||||||   ||| |  ||   |||||
Db        554 GCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGC 613

Qy        655 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGC 714
                |||  |  |||||   |||||   |   |||||  |||||||||  ||| ||| ||| ||
Db        614 TGGCTGCT--GGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGC 671

Qy        715 TAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGGGCT 774
              | |   |||  ||| ||  |||    ||   ||||   ||||  |      |   ||||
Db        672 T-GGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCT 730

Qy        775 GCCGGCCAG 783
              | | | | |
Db        731 GGCTGGCTG 739


RESULT 20
BH284699
LOCUS       BH284699                729 bp    DNA     linear   GSS 30-NOV-2001
DEFINITION  CH230-49O13.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
            CH230-49O13, genomic survey sequence.
ACCESSION   BH284699
VERSION     BH284699.1  GI:17197105
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 729)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 EcoRI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-49O13.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 49  row: O  column: 13
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .729
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-49O13"
```

```
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 1"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      180 a    187 c    150 g    212 t
ORIGIN

  Query Match           12.7%;  Score 127.4;  DB 28;  Length 729;
  Best Local Similarity   77.1%;  Pred. No. 3.8e-19;
  Matches  155;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy        221 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 280
              ||| || |||| || ||  |||| | | || ||||||||||||| ||| || |||||||||
Db        510 GCTAGCCTGCTGGTGGGCTTGCTAGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 569

Qy        281 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 340
              ||| ||| || ||||| |||||| ||| || |||||||||||| ||| || |||||||||
Db        570 GCTTGCTTGCTTGCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 629

Qy        341 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
              ||| |||  | |||||||||||| ||| || ||||||| || | ||| || ||||  |||
Db        630 GCTTGCTTCCTTGCTTGCTTGCTTGCTTGCTTGCCTGATGGCTAGCTTGCTATCTT 689

Qy        401 GCTCGCTGGCCTGCTTGCTGA 421
              ||  ||  | | ||| || |
Db        690 CCTTCCTTACTTACTTTCTTA 710


RESULT 21
BH486116
LOCUS       BH486116                599 bp    DNA     linear   GSS 13-DEC-2001
DEFINITION  BOGJZ09TF BOGJ Brassica oleracea genomic clone BOGJZ09, genomic
            survey sequence.
ACCESSION   BH486116
VERSION     BH486116.1  GI:17694220
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 599)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOGJZ09TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .599
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGJZ09"
                     /clone_lib="BOGJ"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared -
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       63 a    178 c    208 g    150 t
ORIGIN

  Query Match           12.7%;  Score 127;  DB 28;  Length 599;
  Best Local Similarity   53.1%;  Pred. No. 4.4e-19;
  Matches  298;  Conservative    0;  Mismatches  255;  Indels    8;  Gaps    1;

Qy        157 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAG 216
              || |||| | | | |    || |||| | |||||     || ||||  |  |  |
Db         36 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 95

Qy        217 GGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 276
              |  |||| | | ||||    |  |||||| | ||    | || |||||| | ||    |
Db         96 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 155

Qy        277 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 336
              || |||||| | ||    | || |||||| | ||    | || |||||| | ||    |
Db        156 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 215

Qy        337 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 396
              || |||||| | ||    | || |||||| | ||    | || |||||| | ||    |
Db        216 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 275

Qy        397 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACG 456
              || |||||| | ||    | ||       |   |    |  || |||| | |
Db        276 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGT--------C 327

Qy        457 GTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 516
              |       || || | | | | ||     |  |||| ||  ||||    |  |||| ||
Db        328 GCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTC 387

Qy        517 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCT 576
              ||||    |  |||| ||  ||||    |  |||| ||  ||||    |  | || ||
Db        388 GCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTC 447

Qy        577 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCT 636
              ||||    |  |||| ||  ||||    |  | || ||  ||||    |  |||| ||
Db        448 GCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTC 507

Qy        637 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTG 696
              ||||    |  |||| ||  ||||    |  |||| ||  ||||    |  ||| |
Db        508 GCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGACGCTACGTAGCGTGCTCGGAC 567

Qy        697 GCTTGCTTGCTTGCTTGCTAG 717
              |||   | || |||| | | |
Db        568 GCTACGTAGCGTGCTCGGTCG 588


RESULT 22
AZ782118
LOCUS       AZ782118                420 bp    DNA     linear   GSS 16-FEB-2001
DEFINITION  2M0022P01F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0022P01 F, genomic survey sequence.
ACCESSION   AZ782118
VERSION     AZ782118.1  GI:12915490
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 420)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0022  row: P  column: 01
            Seq primer: CGTTGTAAAACGACGGCCAGT
```

```
      Class: plasmid ends
      High quality sequence stop: 420.
FEATURES             Location/Qualifiers
     source          1. .420
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0022P01"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       13 a    219 c     79 g    109 t
ORIGIN

  Query Match             12.7%;  Score 126.6;  DB 28;  Length 420;
  Best Local Similarity   58.7%;  Pred. No. 5e-19;
  Matches  219;  Conservative    0;  Mismatches  154;  Indels    0;  Gaps    0;

Qy         44 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 103
              | | ||| ||   | |  |  ||  | ||||||||||||  ||  |||||| |||
Db         21 TCCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCC 80

Qy        104 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 163
                | |   |||||| | |  ||  |||||| |||      | | |   |||| | |||
Db         81 TCCCTGCCTGCCTGCCTACCTGCCTGCCTGCCTGCATCCCTGCCTGCCTGCCTACCTGCC 140

Qy        164 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 223
                  | | |    ||||| |||  ||  |||||| |||  |   | |     | || |
Db        141 TGCCTGCCTGCCTCCCTGCCTGCCTGCCTGCCTGCCTGCCTACCTGCCTCCCTGCCTCCC 200

Qy        224 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 283
                | | |  |   | ||||  ||  |||||| ||| | |  ||   |||||  | | |||
Db        201 TGCCTACCTGCCTGCCTGCCTGCCTGCCTGCCTCCCTGCCTCCCTGCCTCCCTGCC 260

Qy        284 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 343
                |  |||||| ||| ||| ||  |||||| ||| |||   ||  |||||  | | | |
Db        261 TCCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTTCTATCCTGCCTCCCTCCC 320

Qy        344 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403
                |  |||||| ||| |||  ||  |||| | ||| |||  ||  |||||| ||| |||
Db        321 TCCCTGCCTGCCTGCCTGCCTGCCTACCTGCCTGCCTGCCTGCCTGCCTGCCTGCC 380

Qy        404 CGCTGGCCTGCTT 416
                ||  |||||| |
Db        381 TGCCTGCCTGCCT 393


RESULT 23
CA048466
LOCUS       CA048466                669 bp    mRNA    linear   EST 04-MAR-2003
DEFINITION  ssalsrkc009065 kidney Salmo salar cDNA, mRNA sequence.
ACCESSION   CA048466
VERSION     CA048466.1  GI:24354636
KEYWORDS    EST.
SOURCE      Salmo salar (Atlantic salmon)
  ORGANISM  Salmo salar
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   GRASP Consortium, Davidson,W.S., Koop,B.F. and
            http://web.uvic.ca/cbr/grasp.
  TITLE     A survey of Salmo salar transcripts from high complexity cDNA
            libraries
  JOURNAL   Unpublished
COMMENT     Contact: Koop BF
            Centre for Biomedical Research
            University of Victoria
            PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
            Tel: 250 472 4067
            Fax: 250 472 4075
            Email: bkoop@uvic.ca
            Centre for Biomedical Research, University of Victoria
            cDNA preparation and sequencing:
            Roberto Alberto, Marianne Beetz-Sargent, Maura Busby,
            Peter Hunt, Linda McKinnel, BF Koop.
            bioinformatics:
            Gordon D Brown.
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Salmo salar"
                     /mol_type="mRNA"
                     /strain="McConnell"
                     /db_xref="taxon:8030"
                     /clone_lib="kidney"
                     /note="Vector: pBlueScriptIISK+; Library Creator:  Matthew
                     L Rise ; Atlantic salmon tissue contributors: Carlo Biagi,
                     Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.), Simon
                     Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery (Crofton,
                     B.C.), Rachel Roper (University of Victoria)"
BASE COUNT      112 a    114 c    237 g    206 t
ORIGIN

  Query Match             12.5%;  Score 125.4;  DB 14;  Length 669;
  Best Local Similarity   51.6%;  Pred. No. 1.1e-18;
  Matches  311;  Conservative    0;  Mismatches  291;  Indels    1;  Gaps    1;

Qy         77 GCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTT 136
              ||  |||   ||||  ||| | |||  ||  |   | |  ||| ||| | |  |   ||
Db         10 GCTGGCTGAATGGCAGGCTGGATGACTGGTTAACTGACTGGCTGGCTGGTTAACTGACTG 69

Qy        137 GCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 196
                ||||   | |  | |||| | || |||||  |||     ||    || | |  ||| |
Db         70 GCTGGCTGAATGGCAGGCTTACTGGCTGACTGGATGCCTGGTTGACTGGTTAACTGACTG 129

Qy        197 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 256
                |||  ||||   |    ||       ||||||  |||  || ||   | |  |||  |  |||
Db        130 GCTGGCTGGTTGACTGGTTAACTGACTGGCTGGCTGGTTGGCTGGATAACTGACTCGCTG 189

Qy        257 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 316
                |||  |  |  ||||    ||   |  |  |  |||||   |  |   ||  |  |  |||  |  |  |
Db        190 GCTGGTTGACTGGTTAACTGAGTGGATGGCTGGTTGGTTGACTAGTTAACTGACTGGATG 249

Qy        317 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 376
                |||  |  |  ||||    ||  ||  |  |  |||  |    ||  |  |  ||   |||   |||
Db        250 GCTGGTTGACTGGTTAACTGACTGGATGGCTTGATGACTGGTTAACTGAGTGGATGGCTG 309

Qy        377 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 436
                |  |  |  ||||    ||  ||  |||  |||||  ||  ||  |  |   |||  |  ||||
Db        310 GTTGGTTGACTGGTTACCTGACTGGCTGGCTGG-TTGACTGGTTAACTGACTGGATGGCT 368

Qy        437 GCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCT 496
                | |||  ||  |  |   |||  |  |  |  |||  |||   |  ||  |   |   |||||   ||  |
Db        369 GGTTGACTGGTTAACTGACTGGATGGCTGGATGACTGGTTCACTGACTGGCTGGCTGGTT 428

Qy        497 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 556
```

```
Db        429 GACTGGTTCACTGACTGGCTGGCTGGTTAACTGACTGGCTGGCTGAATGGATGGCTGGTT 488

Qy        557 GCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCT 616

Db        489 GACTGGTTAACTGACTGGCTGGCTGGTTGACTGGTTAACTGACTGGATGGCTGGTTGACT 548

Qy        617 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 676

Db        549 GGTTAACTGACTGGATGGCTGGTTGACTGGTTAACTGACTGGATGGCTGGATGACTGGTA 608

Qy        677 GCT 679

Db        609 ACT 611
```

RESULT 24
BZ041698
LOCUS BZ041698 729 bp DNA linear GSS 09-OCT-2002
DEFINITION llr23h12.b1 B.oleracea002 Brassica oleracea genomic, genomic survey sequence.
ACCESSION BZ041698
VERSION BZ041698.1 GI:23631988
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 729)
AUTHORS Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash ,W., Rabinowicz,P.D. and Wilson,R.K.
TITLE Whole genome shotgun reads from Brassica oleracea
JOURNAL Unpublished
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Plate: llr23 row: h column: 12
Seq primer: -21UPpOT forward
Class: shotgun
High quality sequence start: 8
High quality sequence stop: 551.
FEATURES Location/Qualifiers
source 1. .729
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pOTw13; Whole genome shotgun library from flowering buds. DNA was purified from a crude nuclear prep using Brassica oleracea TO1000DH3 buds provided by Thosmas Osborn at the University of Wisconsin. Genomic DNA was provided by Pablo Rabinowicz (CSHL) and the shotgun library prepared at Washington University Genome Sequencing Center."
BASE COUNT 114 a 237 c 235 g 141 t 2 others
ORIGIN

Query Match 12.4%; Score 124.4; DB 28; Length 729;
Best Local Similarity 53.0%; Pred. No. 1.9e-18;
Matches 289; Conservative 0; Mismatches 252; Indels 4; Gaps 1;

```
Qy         70 GCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGG 129

Db        187 GCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTAC 246

Qy        130 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCG 189

Db        247 GCAGCGTGCTTGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTAC 306

Qy        190 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGG 249

Db        307 GCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCCCGGTCGCTAC 366

Qy        250 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 309

Db        367 GCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTAC 426

Qy        310 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 369

Db        427 GCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTAC 486

Qy        370 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 429

Db        487 GCAGCGTGCCCGGTCGCTATGCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTAC 546

Qy        430 TAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTG 489

Db        547 GCAGCATGCCCGGTCGCTACGCACATGCT----CAGCGTGCTCGGTCGCTACGCANCGTG 602

Qy        490 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 549

Db        603 CTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTG 662

Qy        550 CTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTC 609

Db        663 CTCGGTCGCTACGCAGAGTGCTCGGTCGCTACGCAGCGTGGTCCGGCGCTACGACGCGTG 722

Qy        610 CTGGG 614

Db        723 CTCGG 727
```

RESULT 25
BH604586
LOCUS BH604586 633 bp DNA linear GSS 15-DEC-2001
DEFINITION BOHCF82TF BOHC Brassica oleracea genomic clone BOHCF82, genomic survey sequence.
ACCESSION BH604586
VERSION BH604586.1 GI:17857032
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 633)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished
COMMENT Other_GSSs: BOHCF82TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .633
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOHCF82"
/clone_lib="BOHC"
/note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 80 a 207 c 212 g 134 t
ORIGIN

Query Match 12.3%; Score 123.4; DB 28; Length 633;
Best Local Similarity 52.7%; Pred. No. 3e-18;

```
  Matches  321; .Conservative    0;  Mismatches  276;  Indels   12;  Gaps    2;

Qy       113 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACA 172
Db         7 GCGTGCTCGGTCCCTACGCAGCGTGCTCGGTCCCTACGCAGCGTGCTCGGTCCCTACGCA 66

Qy       173 GTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 232
Db        67 GCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCACTACCCAGCGTGCTCGGTCGCTACGCA 126

Qy       233 GTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 292
Db       127 GCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCA 186

Qy       293 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 352
Db       187 GCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCA 246

Qy       353 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 412
Db       247 GCGTGTTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCA 306

Qy       413 GCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATG 472
Db       307 GCGTGCTCGGTCGCAACGCAGCGTGCTTGGTCGCTACGTAGCGTGCTCGGTCCCC----- 361

Qy       473 GCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 532
Db       362 ------TACGTAGCGTGCTCGGTCCCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTC 415

Qy       533 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTG 592
Db       416 GGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCAACGTAGCGTGCTC 475

Qy       593 GGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 652
Db       476 GGTCGCAACGTAGCGTGCTCGGTTGCTATGTAGCGTGCACGGCGCTACGTAGCACGTAGC 535

Qy       653 GGCTGCTAG-TAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCT 711
Db       536 GACCGAGCGATCGTCCCGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCT 595

Qy       712 TGCTAGTGG 720
Db       596 CGGTCGCGG 604


RESULT 26
AZ633613/c
LOCUS       AZ633613                591 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  1M0488O23R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0488O23 R, genomic survey sequence.
ACCESSION   AZ633613
VERSION     AZ633613.1  GI:11755803
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0488  row: O  column: 23
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 591.
FEATURES             Location/Qualifiers
     source          1. .591
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0488O23"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      243 a    201 c     61 g     86 t
ORIGIN

  Query Match             12.2%;  Score 122.4;  DB 28;  Length 591;
  Best Local Similarity   52.6%;  Pred. No. 5.1e-18;
  Matches  267;  Conservative    0;  Mismatches  241;  Indels    0;  Gaps    0;

Qy       228 TGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 287
Db       510 TGCTTGGTTGGTTGGGTGCTTGCTTGTTTGGTTGCTTACATGCTTGCTTGCTTGCTTGCT 451

Qy       288 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 347
Db       450 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTAGATGATTAGTTGGCTCATTGGTTTGTTGGT 391

Qy       348 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 407
Db       390 TCATTGGTCAGTCGGTTGGTCGGTCGGTTGGTCAGTCATTTGATTGGTTGGATGGTTGGT 331

Qy       408 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCC 467
Db       330 TAGTTGGTTAGTTGGCTGAGTGATTGCTTGGGTGATTGACTGAAAGACTGGTTGTTTGGT 271

Qy       468 TGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 527
Db       270 TGCTTGGTTGATTGCTTGGTTGGGTGACTGGTTGGTTGGGTGACTGGCTGGTTGGTTGGT 211

Qy       528 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGC 587
Db       210 TGTTTGGTTGGTTGTTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGGTGAGTGAGTGAG 151

Qy       588 TGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 647
Db       150 TCAGTGAGTGAGTGAGTGAGTGACTGCTTGGTTAGGCAATTGGTTGGTAAATTGTTTAGT 91

Qy       648 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCT 707
```

```
Db         90 TGGGTTGTTGGTTGAATGCTTCCTTGAGTAATTGGCTGGAAGATTGGTTGCTTGATTGCT 31

Qy        708 TGCTTGCTAGTGGGGCCGCTTGCCTGCT 735
              || ||  | |   ||  |||||| ||||
Db         30 TGGTTAGTTGGTTGGTTGCTTGCTTGCT 3


RESULT 27
AK028099/c
LOCUS       AK028099               1142 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus adult male testis cDNA, RIKEN full-length enriched
            library, clone:1700125N06 product:unclassifiable, full insert
            sequence.
ACCESSION   AK028099
VERSION     AK028099.1  GI:26080728
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 1142)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .1142
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:1700125N06"
                     /db_xref="taxon:10090"
                     /clone="1700125N06"
                     /sex="male"
                     /tissue_type="testis"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     misc_feature    1. .1142
                     /note="unclassifiable"
BASE COUNT      303 a    340 c    271 g    228 t
ORIGIN

  Query Match             12.2%;  Score 122.4;  DB 11;  Length 1142;
  Best Local Similarity   72.3%;  Pred. No. 6e-18;
  Matches  159;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy        476 TGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGC 535
              || | |    |||||||||||||| |   |||||||||||||| |    | ||||| ||
Db        436 TGCTGGGCAAGCTGCTGGGCTGCTGGGCAGGCTGCTGGGCTGCTGGGCAAGTTGCTGAGC 377

Qy        536 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGC 595
              ||||  |   |||||||||||||| |   || | |||||||||| |   ||||||||||
Db        376 TGCTGGGCAGGCTGCTGGGCTGCTGGACAGGATGCTGGGCTGCTGGGCAGGCTGCTGGGC 317

Qy        596 TGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGC 655
              |||| | | || | |||||||||| |   || |||||| |||| |   || |||||||
Db        316 TGCTGGGAAGGTTGCTGGGCTGCTGGGCAGGGTGCTGGGATGCTGGGTAGGGTGCTGGGA 257

Qy        656 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCT 695
              |||| |   || |||||| |||| |   || ||| || |
Db        256 TGCTGGGTAGGGTGCTGGGATGCTGGGTAGGGTGCTTGGT 217


RESULT 28
AZ650539
```

```
LOCUS       AZ650539               693 bp    DNA     linear   GSS 14-DEC-2000
DEFINITION  1M0520I14R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0520I14 R, genomic survey sequence.
ACCESSION   AZ650539
VERSION     AZ650539.1  GI:11785128
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 693)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0520  row: I  column: 14
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 693.
FEATURES             Location/Qualifiers
     source          1. .693
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0520I14"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      144 a    156 c    166 g    226 t      1 others
ORIGIN

  Query Match           12.2%;  Score 122;  DB 28;  Length 693;
  Best Local Similarity   68.0%;  Pred. No. 6.6e-18;
  Matches  170;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy        170 ACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTG 229
              | | ||  |  |||  ||     || |   |  | |  | |  |   |  |||| |||||
Db        439 AGACTATACAGGCTAACTGCTAAGCTTTGCTTTTTAATGAAAGGCTTGCTTGCTTGCTTG 498

Qy        230 CTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 289
              |  |   | |||||  ||  |||||| |||  |||| ||  ||  ||  |||||||  ||| |
Db        499 CCTGCTTGCCTGCTTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGTCTGCTTGCCTGCTTG 558

Qy        290 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 349
              |||||||||||||| ||| || ||||||||||||| ||| ||  ||||||||||  ||  |
Db        559 CCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTAGCTTGCTTGCCTGCCTG 618

Qy        350 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 409
              ||||| ||| |||  ||  |||||| ||| ||   ||| | ||||||| |||| ||  |
Db        619 CCTGCCTGCCTGCCTGCCTGCCTGCCTGACTGCTTGTCTGCTTGCCTGCTTGCCTG 678

Qy        410 CCTGCTTGCT 419
              | ||||||||
Db        679 CTTGCTTGCT 688


RESULT 29
BZ015565
LOCUS       BZ015565               726 bp    DNA     linear   GSS 07-OCT-2002
DEFINITION  oeh87c12.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
            sequence.
ACCESSION   BZ015565
VERSION     BZ015565.1  GI:23572725
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 726)
  AUTHORS   Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash
            ,W., Rabinowicz,P.D. and Wilson,R.K.
  TITLE     Whole genome shotgun reads from Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: oeh87  row: c  column: 12
            Seq primer: -21UPpOT forward
            Class: shotgun
            High quality sequence start: 45
            High quality sequence stop: 398.
FEATURES             Location/Qualifiers
     source          1. .726
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3712"
                     /clone_lib="B.oleracea002"
                     /note="Vector: pOTw13; Whole genome shotgun library from
                     flowering buds. DNA was purified  from a crude nuclear
                     prep using Brassica oleracea TO1000DH3 buds  provided by
                     Thosmas Osborn at the University of Wisconsin. Genomic
                     DNA was provided by Pablo Rabinowicz (CSHL) and the
                     shotgun library  prepared at Washington University Genome
                     Sequencing Center."
BASE COUNT       51 a    250 c    271 g    153 t      1 others
ORIGIN

  Query Match           12.0%;  Score 120.2;  DB 28;  Length 726;
  Best Local Similarity   53.1%;  Pred. No. 1.7e-17;
  Matches  280;  Conservative    0;  Mismatches  243;  Indels    4;  Gaps    1;

Qy        228 TGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 287
              | ||||   |   | |||||    || |||| |  |||||    || |||| |  ||||
Db         35 TTCTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCCGCT 94

Qy        288 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 347
              |    || |||| |  |||||    || |||| |  |||||    || |||| |  ||||
Db         95 GCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCCGCT 154

Qy        348 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 407
              |    || |||| |  | |||    || |||| |  |||||    || |||| |  ||||
Db        155 GCGTAGCGTGCTCGGCCACTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCCGCT 214
```

```
Qy        408 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCC 467
Db        215 GCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCAGCT 274

Qy        468 TGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 527
Db        275 GCGTAGCGTGCT----CGGCCGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGC 330

Qy        528 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGC 587
Db        331 CGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGC 390

Qy        588 TGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 647
Db        391 CGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGC 450

Qy        648 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCT 707
Db        451 CGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGC 510

Qy        708 TGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGC 754
Db        511 CGCTGCGTAGCCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCCGCTGC 557


RESULT 30
BB610246/c
LOCUS       BB610246                659 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB610246 RIKEN full-length enriched, adult male testis Mus musculus
            cDNA clone 1700125N06 5', mRNA sequence.
ACCESSION   BB610246
VERSION     BB610246.1  GI:16451804
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 659)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES              Location/Qualifiers
     source           1. .659
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="1700125N06"
                      /sex="male"
                      /tissue_type="testis"
                      /dev_stage="adult"
                      /lab_host="SOLR"
                      /clone_lib="RIKEN full-length enriched, adult male testis"
                      /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
                      prepared and sequenced in Mouse Genome Encyclopedia
                      Project of Genome Exploration Research Group in Riken
                      Genomic Sciences Center and Genome Science Laboratory in
                      RIKEN. Division of Experimental Animal Research in Riken
                      contributed to prepare mouse tissues. 1st strand cDNA was
                      primed with a primer [5'
                      GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                      prepared by using trehalose thermo-activated reverse
                      transcriptase and subsequently enriched for full-length by
                      cap-trapper. Second strand cDNA was prepared with the
                      primer adapter of sequence [5'
                      GAGAGAGAGAGCGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC
                      3']. cDNA was cloned into the XhoI and BamHI sites. "
BASE COUNT      158 a    229 c    146 g    125 t      1 others
ORIGIN

  Query Match             12.0%;  Score 119.8;  DB 10;  Length 659;
  Best Local Similarity   71.4%;  Pred. No. 2.1e-17;
  Matches  157;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        476 TGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGC 535
Db        433 TGCTGGGCAAGCTGCTGGGNTGCTGGGCAGGCTGCTGGGCTGCTGGGCAAGTTGCTGAGC 374

Qy        536 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGC 595
Db        373 TGCTGGGCAGGCTGCTGGGCTGCTGGACAGGATGCTGGGCTGCTGGGCAGGCTGCTGGGC 314

Qy        596 TGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGC 655
Db        313 TGCTGGGAAGGTTGCTGGGGTGCTGGGCAGGGTGCTGGGATGCTGGGTAGGGTGCTGGGA 254

Qy        656 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCT 695
Db        253 TGCTGGGTAGGGTGCTGGGATGCTGGGTAGGGTGCTTGGT 214


RESULT 31
AL528199
LOCUS       AL528199                1100 bp    mRNA    linear   EST 23-MAY-2003
DEFINITION  AL528199 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DC025YF13 5-PRIME, mRNA sequence.
ACCESSION   AL528199
VERSION     AL528199.2  GI:31066049
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1100)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
```

```
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 13, 2001 this sequence version replaced gi:12791692.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 1419.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DC025CC07QP1&cluster=1419.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DC025CC07QP1.
FEATURES              Location/Qualifiers
     source           1. .1100
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CS0DC025YF13"
                      /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
                      /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
                      /note="1st strand cDNA was primed with a NotI-oligo(dT)
                      primer. Five prime end enriched, double-strand cDNA was
                      digested with Not I and cloned into the Not I and EcoR V
                      sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT       74 a    296 c    463 g    258 t      9 others
ORIGIN

  Query Match             12.0%;  Score 119.6;  DB 9;  Length 1100;
  Best Local Similarity   61.9%;  Pred. No. 2.6e-17;
  Matches  394;  Conservative    0;  Mismatches  219;  Indels   23;  Gaps   12;

Qy        112 GGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGAC 171
              ||| ||| ||| ||||||  ||| ||||  ||  | |   | |||  ||    |  | |
Db        251 GGCTGGCTGGCTGGCTGGCTGGCTGCTGG--CTGGCTGGCTGGCTGGCTGGCTGGC 308

Qy        172 AGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG------GACAGTAGGGCTGCTGG 225
              |  |||  ||| ||| ||||||  ||| ||||  ||      | | |   |||||   |
Db        309 TGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTG 368

Qy        226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
              ||||  |  ||    ||| |||||  ||| || |||| ||||||  ||| ||| ||| |
Db        369 GCTGCTGGCTGGCTGGCTGGCTGGCTGGC-TGCTGGCTGGCTGGCTGGCTGGCTGG 427

Qy        286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
              |||| ||||| ||| ||| ||||||  ||| ||| ||| ||||||  ||| ||| ||| |
Db        428 CTGG-CTGCTGGCTGGCTGGCTGGCTGGCTGGCT-GCTGGCTGGCTGGCTGGCTGG 485

Qy        346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 405
              |||| |||  |||| ||| ||||||  ||| ||| ||| ||| ||  ||| ||| ||| |
Db        486 CTGG-CTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCT-GCTGGCTGGCTGGCTGG 543

Qy        406 CTGGCCTGCTTGCTGAGGG--GACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCC 463
              |||||  ||| ||||  ||  | | |   |||||  ||  ||       ||  ||   ||
Db        544 CTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCT 603

Qy        464 TGCCTGATGGCTTGATAGTAG--GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAG 521
              ||  | ||||| | |  | |  |||||   |||||   |  |||||||||   ||| |
Db        604 GGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGG 663

Qy        522 TAGGGCTGCTGGGCTGCTAGTAG--GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCT 579
              ||   |||||   |||  | |  |||||   |||||   |  .||||  ||||   |||
Db        664 CTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCT 723

Qy        580 AGTAGGGCTGCTGGGCTGCTAGTAG--GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTG 637
              |  ||   |||||   |||  | |  |||   |||||   |   |||||||||   |
Db        724 GGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGG 783

Qy        638 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGG 697
              || |  ||   |||||   |||  |  |||||   |||||   |   |||||  |||||||
Db        784 CTGGCTGGCTGGCTGGCTGGCTGCT--GGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGG 841

Qy        698 CTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTG 733
                  || ||| ||| ||| ||| |  ||     | || ||
Db        842 CTGGCTGGCTGGCTGGCTGGCTGGCTGCATGGCATG 877


RESULT 32
BZ236616
LOCUS       BZ236616                736 bp    DNA     linear   GSS 12-OCT-2002
DEFINITION  CH230-251M10.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-251M10, genomic survey sequence.
ACCESSION   BZ236616
VERSION     BZ236616.1  GI:23895644
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 736)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-251M10.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 251  row: M  column: 10
            Seq primer: T7
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .736
                      /organism="Rattus norvegicus"
                      /mol_type="genomic DNA"
                      /strain="BN/SsNHsd/MCW"
                      /db_xref="taxon:10116"
                      /clone="CH230-251M10"
                      /sex="Female"
                      /cell_type="Brain"
                      /clone_lib="CHORI-230 Segment 2"
                      /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                      CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                      Pieter de Jong"
BASE COUNT      143 a    168 c    196 g    229 t
ORIGIN

  Query Match             11.9%;  Score 118.6;  DB 29;  Length 736;
  Best Local Similarity   69.9%;  Pred. No. 4e-17;
  Matches  160;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy        191 TGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGC 250
              | ||  || |||||  |||   |   || |||| |||||||     |  |||| ||| ||
Db         14 TAGCTAGCCTGCTGGTGGGCTTGCTAGGTTGCTTGCTTGCTTCCTTGCTTGCTTGCTTGC 73

Qy        251 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 310
              ||||||||||||  ||| || |||||||| ||  ||  || | |||||| || |||  |
Db         74 TTGCTTGCTTGCTTGCTTGCTTGCTTTCTTCCTTGCTTCCTTGCTTCCTTGCTTCC 133

Qy        311 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 370
              |||||  ||||||  || || | |||||| || |||  | |||||  ||||||  || ||
Db        134 TTGCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTTGC 193
```

```
Qy        371 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              ||||||| || | ||| || |||| ||| ||  ||  | | ||| ||
Db        194 TTGCTTGCCTGATGGCTAGCTTGCTATCTTCCTTCCTTCCTTCCTTCCT 242


RESULT 33
AZ608625/c
LOCUS       AZ608625                680 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  1M0433A01F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0433A01 F, genomic survey sequence.
ACCESSION   AZ608625
VERSION     AZ608625.1  GI:11730815
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 680)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0433  row: A  column: 01
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 680.
FEATURES             Location/Qualifiers
     source          1. .680
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0433A01"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT     253 a    223 c    113 g     91 t
ORIGIN

  Query Match            11.8%;  Score 118;  DB 28;  Length 680;
  Best Local Similarity  59.2%;  Pred. No. 5.4e-17;
  Matches  219;  Conservative    0;  Mismatches  150;  Indels    1;  Gaps    1;

Qy        199 TTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGC 258
              |||||||   |    | ||| |   |  | | | |   |||||   | |||  || |||
Db        498 TTGCTGATCAGTGGTTTGGGTTTGAGTTTGGGTTGGTTGGCTGGCTGGTGGGTTGGTTGG 439

Qy        259 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 318
              ||||   | || || | | | | | ||||| ||| ||  ||||| ||| || |||||||
Db        438 CTGCTTAGTTGCTTGTTGGTTGGTTGGCTGG-CTGGTTAGTTGCTTGCTTGCTTGCTTGC 380

Qy        319 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 378
              ||||| ||| || ||||||||||||| ||| || |||||||||||| ||| || ||
Db        379 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGGCAAG 320

Qy        379 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGC 438
              ||| | | | | |||  || ||| | |   | |||  || |  |  |  |   |||||
Db        319 TTGGTTGGTTGACTGGCTGGTTGGTTGGCTGGCTGGCTGGTTGGTTGGTTGGTTGGCTGG 260

Qy        439 TTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGC 498
              ||| | | | ||  | | |   |  ||  || | | ||  | |   | ||| | || |
Db        259 TTGGTAGGTTAGTTGGCTGGTTGGTTGGTTGGTTGGTTAGTTGGTTGACTGGTTGGTTAG 200

Qy        499 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 558
              | |   | ||| | || || | |   || || | || || | |   || || | || ||
Db        199 TTGGTTGACTGGTTGGTTGGTTGACTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGG 140

Qy        559 TAGTAGGGCT 568
              | |   || |
Db        139 TTGGTTGGTT 130


RESULT 34
CA047817
LOCUS       CA047817                712 bp    mRNA    linear   EST 04-MAR-2003
DEFINITION  ssalbrh012052 head Salmo salar cDNA, mRNA sequence.
ACCESSION   CA047817
VERSION     CA047817.1  GI:24354003
KEYWORDS    EST.
SOURCE      Salmo salar (Atlantic salmon)
  ORGANISM  Salmo salar
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
REFERENCE   1  (bases 1 to 712)
  AUTHORS   GRASP Consortium, Davidson,W.S., Koop,B.F. and
            http://web.uvic.ca/cbr/grasp.
  TITLE     A survey of Salmo salar transcripts from high complexity cDNA
            libraries
  JOURNAL   Unpublished
COMMENT     Contact: Koop BF
            Centre for Biomedical Research
            University of Victoria
            PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
            Tel: 250 472 4067
            Fax: 250 472 4075
            Email: bkoop@uvic.ca
            Centre for Biomedical Research, University of Victoria
            cDNA preparation and sequencing:
            Roberto Alberto, Marianne Beetz-Sargent, Maura Busby,
            Peter Hunt, Linda McKinnel, BF Koop.
            bioinformatics:
            Gordon D Brown
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .712
                     /organism="Salmo salar"
                     /mol_type="mRNA"
                     /strain="McConnell"
                     /db_xref="taxon:8030"
                     /clone_lib="head"
                     /note="Vector: pBlueScriptIISK+; Library Creator:
                     Kristian R von Schalburg ; Atlantic salmon tissue
                     contributors: Carlo Biagi, Mitch Uh and Robert Devlin (DFO
```

```
                , Vancouver, B.C.), Simon Jones (PBS, Nanaimo, B.C.),
                Seaspring Hatchery (Crofton, B.C.), Rachel Roper
                (University of Victoria)"
BASE COUNT      168 a    159 c    194 g    191 t
ORIGIN

  Query Match             11.8%;  Score 118;  DB 14;  Length 712;
  Best Local Similarity   58.6%;  Pred. No. 5.5e-17;
  Matches  205;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy          72 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 131
               | |  ||  ||| |||| |  ||| |||||  ||   | ||||  ||| ||| ||||||
Db         337 TACTGGCTGGCTGGCTGACTGGCTGGCTGACTGGCTGGCAGGCTGGCTGGCTGGCTGGCT 396

Qy         132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191
                 ||  |||   |    |  |||   ||  ||||  ||   | ||||  ||  ||| |||
Db         397 GACTAACTGGCTGACTGGCTGGCTGACTGACTGACTGGCTGGCAGGCTGGCAGGCTGGCT 456

Qy         192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251
               |||  |||  |||   |       ||    ||| ||  || |  |     || |||||||
Db         457 GGCTGGCTGACTGGCTGACTGACTGGCTGACTGACTGACTGGCTGACTGACTGGCTGGCT 516

Qy         252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
               |||  ||  || ||||||  |||  ||  ||  ||| |   || |||  ||  |||||
Db         517 GGCTGACTGACTGGCTGGCTGGCTGACTAACTGACTGACTGACTGGCTGACTAACTGGCT 576

Qy         312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
               ||| ||| ||| ||||||  |||  ||  ||  ||| |  ||| |||  ||  |||||
Db         577 GGCTGGCTGGCTGGCTGGCTGGCTGACTAACTGACTGACTGGCTGGCTGACTGACTGGCT 636

Qy         372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGA 421
               |||  ||  || ||||||  |||  ||  || ||||||   ||  || |
Db         637 GGCTGACTGACTGGCTGGCTGGCTGACTGACTGGCTGGCTGACTGACTAA 686


RESULT 35
BZ285588
LOCUS       BZ285588                549 bp    DNA     linear   GSS 15-OCT-2002
DEFINITION  CH230-385O23.TJC CHORI-230 Segment 2 Rattus norvegicus genomic
            clone CH230-385O23, genomic survey sequence.
ACCESSION   BZ285588
VERSION     BZ285588.1  GI:24016713
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-385O23.TVB
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 385  row: O  column: 23
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-385O23"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      149 a    104 c    211 g     85 t
ORIGIN

  Query Match             11.7%;  Score 117.2;  DB 29;  Length 549;
  Best Local Similarity   53.7%;  Pred. No. 7.9e-17;
  Matches  288;  Conservative    0;  Mismatches  243;  Indels    5;  Gaps    2;

Qy         149 GTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCT-TGCTCGCTGGCCTGCTTGCTGAGG 207
               | | ||  || || | || ||   | |||||||  |   |   | ||    ||| | | ||
Db          14 GGATGCAAGCCTGATTAGAGGGGTGCAGGCCTGATCAGAGGGATGCAAGGCTGGATCAGA 73

Qy         208 GGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCT 267
               ||   | | | |||        |   ||| | ||| ||   ||  |||  || || ||
Db          74 GGGATGCAAGCCTGACCAGAGGGGTGTAAGCCTGATCAGAGGGGTGCAAGCCTGATCAGA 133

Qy         268 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 327
               ||  |||  |  || ||   ||  |||  || || ||   ||  |||  || || ||
Db         134 GGGATGCAAGGCTGATCAGAGGGATGCAAGCCTGATCAGAGGGGTGCAAGCCTGATCAGA 193

Qy         328 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 387
               ||  |||  || || ||   ||  |||  || || ||   ||  |||  |  || ||
Db         194 GGGGTGCAAGCCTGATCAGAGGGATGCAAGCCTGATCAGAGGGATGCAAGGCTGATCAGA 253

Qy         388 GGCCTGCTTGCTTGCTC----GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCT 443
               ||  |||  || || ||    |  ||  || || | || ||   | | | |||  |
Db         254 GGGATGCAAGCCTGATCAGAGGGATGCAAGCCTGATCAGAGGGGTGCAAGCCTGATCAGA 313

Qy         444 TGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTA 503
               |      | | |  |    |  ||   || ||||  |||| ||   | ||| |   |
Db         314 GGGGTGCAAGCCTGATCAGAGGGGTGCAAGCCTGATCAGAGGGATGAAAGCCTGATCAGA 373

Qy         504 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 563
               ||| |||  | ||| |   |||| |||  ||||| |   |||| |||  ||||| |   |
Db         374 GGGGTGCAAGCCTGATCAGAGGGATGCAAGGCTGATCAGAGGGGTGCAAGGCTGATCAGA 433

Qy         564 GGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTA 623
               ||| | |  | ||| |   |||| |||  | ||| |   |||| | |  | ||| |   |
Db         434 GGGATGCAAGCCTGATCAGAGGGGTGCAAGCCTGATCAGAGGGGTGCAAGCCTGATCAGA 493

Qy         624 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 679
               ||| |||  | ||| |   |||| ||   | ||| |   |||| |||  | ||| |
Db         494 GGGGTGCAAGCCTGATCAGAGGGATGAAAGCCTGATCAGAGGGGTGCAAGCCTGAT 549


RESULT 36
BH328866
LOCUS       BH328866                740 bp    DNA     linear   GSS 03-DEC-2001
DEFINITION  CH230-93P12.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
            CH230-93P12, genomic survey sequence.
ACCESSION   BH328866
VERSION     BH328866.1  GI:17259580
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 740)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 EcoRI segment
```

```
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-93P12.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 93  row: P  column: 12
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .740
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-93P12"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 1"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      181 a    199 c    149 g    211 t
ORIGIN

  Query Match             11.7%;  Score 117;  DB 28;  Length 740;
  Best Local Similarity   76.2%;  Pred. No. 9.5e-17;
  Matches  144;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy        231 TAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 290
              |||   | ||||| |  ||| |||| | ||||| ||| || ||||||||||||| ||| ||
Db        522 TAGCTAGCCTGCTGGTGGGCTTGCTAGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 581

Qy        291 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 350
              | ||| |||||| ||| || |||||||||||| |||  | ||||||||||||| ||| ||
Db        582 TTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTCCTTGCTTGCTTGCTTGCTTGC 641

Qy        351 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 410
              | |||||||||| |||  | ||||| |||||| ||| ||||| | ||| ||| |||  |
Db        642 TTTCTTGCTTGCTTGCTTCCTTGCTTCCTTGCTTGCTTGCCTGATGGCTAGCTTGCTATC 701

Qy        411 CTGCTTGCT 419
              | ||| ||
Db        702 TTCCTTCCT 710


RESULT 37
BH515818
LOCUS       BH515818                564 bp    DNA     linear   GSS 13-DEC-2001
DEFINITION  BOHAC43TR BOHA Brassica oleracea genomic clone BOHAC43, genomic
            survey sequence.
ACCESSION   BH515818
VERSION     BH515818.1  GI:17723908
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 564)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOHAC43TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .564
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHAC43"
                     /clone_lib="BOHA"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       56 a    176 c    198 g    134 t
ORIGIN

  Query Match             11.6%;  Score 115.8;  DB 28;  Length 564;
  Best Local Similarity   51.9%;  Pred. No. 1.7e-16;
  Matches  291;  Conservative    0;  Mismatches  262;  Indels    8;  Gaps    1;

Qy        157 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAG 216
              || ||||  |  |  |    || |||| | |||||     || ||||  |  |  |
Db          1 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 60

Qy        217 GGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 276
              |  |||| | | ||||    |  |||||| | ||    | || ||||||| | ||    |
Db         61 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 120

Qy        277 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 336
              || |||||| | ||    | || |||||| | ||    | || |||||| | ||    |
Db        121 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 180

Qy        337 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 396
              || |||||| | ||    | || |||||| | ||    | || |||||| | ||    |
Db        181 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 240

Qy        397 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACG 456
              || |||||| | ||    | ||      |  |    |  || |||| | |
Db        241 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGT--------C 292

Qy        457 GTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 516
              |       ||  | | | | ||    |  ||| ||  ||||    |  ||| ||
Db        293 GCTACGTAGCGCGCTCGGTCGCTACGTAGCGCGCTCGGTCGCTACGTAGCGCGCTCGGTC 352

Qy        517 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCT 576
              ||||    |  | | ||  ||||    |  ||| ||  ||||    |  || ||
Db        353 GCTACGTAGCGCGTTCGGTCGCTACGTAGCGCGCTCGGTCGCTACGTAGCGCGCTCGGTC 412

Qy        577 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCT 636
              ||||    |  ||| ||  ||||    |  | || ||  ||||    |  |||| ||
Db        413 GCTACGTAGCGCGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTC 472

Qy        637 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTG 696
              ||||    |  |||| ||  ||||    |  |||| ||  ||||    |  ||| | |
Db        473 GCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTC 532

Qy        697 GCTTGCTTGCTTGCTTGCTAG 717
              |||   | || |||| | | |
Db        533 GCTACGTAGCGTGCTCGGTCG 553


RESULT 38
BM950444
LOCUS       BM950444                711 bp    mRNA    linear   EST 14-MAR-2002
DEFINITION  UI-M-EH0p-buu-i-05-0-UI.r1 NIH_BMAP_EH0p Mus musculus cDNA clone
            IMAGE:5687116 5', mRNA sequence.
ACCESSION   BM950444
```

```
VERSION     BM950444.1  GI:19434034
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 711)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin, Univeristy of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .711
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5687116"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 18.5 dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_EH0p"
                     /note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured mRNa was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site. Double stranded
                     cDNA was size selected according to mRNA size fraction,
                     ligated with EcoR I adaptor, digested with Not I, and then
                     cloned directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     , is CAGCCACGAC. This library was created for the
                     University of Iowa Mouse Brain Molecular Anatomy Project
                     (BMAP): 'Gene Discovery in the Developing Mouse Nervous
                     System', supported by National Institute of Mental Health
                     (NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT      177 a    155 c    271 g    108 t
ORIGIN
```

```
  Query Match           11.6%;  Score 115.6;  DB 12;  Length 711;
  Best Local Similarity   52.2%;  Pred. No. 2e-16;
  Matches  282;  Conservative    0;  Mismatches  254;  Indels    4;  Gaps    1;

Qy        155 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 214
              ||||||    ||     | |||    |  | |    ||  |   ||   | | |
Db         42 CTGCTTATGTAGCACCATGCAGGGAAACCAGGTGTGCCGCAGGGAAGCCAGGTGTGCTTC 101

Qy        215 AGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 274
              ||||  ||  | |     | ||||  ||  | ||  ||||  |   ||  | ||  ||||
Db        102 AGGGAAGCCAGGTGTACTGCAGGGAAGCCAGGTGTACTGCAGGGAAGCCAGGTGTGCTGC 161

Qy        275 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 334
                |   ||  | ||  ||||  |   ||  | ||  ||||  |   ||  | ||  | ||
Db        162 AGGGAAGCCAGGTGTACTGCAGGGAAGCCAGGTGTACTGCAGGGAAGCCAGGTGTGCAGC 221

Qy        335 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 394
                |   ||  | ||  || |  |   ||  | ||  ||||  |   ||  | ||  ||||
Db        222 AGGGAAGCCAGGTGTGCTTCAGGGAAGCCAGGTGTACTGCAGGGAAGCCAGGTGTGCTGC 281

Qy        395 TTGCTTGCTCGCTGGCCTGCTT----GCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAG 450
                |   ||  | ||  ||||      ||   | | || | ||||  ||  | |    |
```

```
Db        282 AGGGAAGCCAGTTGTGCTGCAGGGAAGCCAGGTGTACTGCAGGGAAGCCAGGTGTGCAGC 341

Qy        451 GGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 510
              ||   |   |     |||  ||   |  ||  | ||||| |||   | || || |||||
Db        342 AGGGAAGCCAGGTGTGCTGCAGGGAAGCCAGGTGTGCTGCAGGGAAACCAGGAGTGCTGC 401

Qy        511 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCC 570
              |||  || ||  | ||||| |||   | || || |||||  ||   || ||  | ||| |
Db        402 AGGGAAGCCAGGTGTGCTGCAGGGAAACCAGGAGCGCTGCAGGTAAGCCAGGTGCGCTGC 461

Qy        571 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 630
              |||  || ||  | ||||| |||  || ||  | ||| | |||  || ||  | |||||
Db        462 AGGGAAGCCAGGTGTGCTGCAGGGAAGCCAGGTGTGCTGCAGGGAAGCCAGGTGTGCTGC 521

Qy        631 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 690
              ||   || ||  | ||||| |||  || ||  | |||||  ||   || ||  | |||||
Db        522 AGGAAAGCCAGGTGCGCTGCAGGGAAGCCAGGTGTGCTGCAGGAAAGCCAGGTGTGCTGC 581
```

```
RESULT 39
BH443125
LOCUS       BH443125                490 bp    DNA     linear   GSS 12-DEC-2001
DEFINITION  BOHRJ25TR BOHR Brassica oleracea genomic clone BOHRJ25, genomic
            survey sequence.
ACCESSION   BH443125
VERSION     BH443125.1  GI:17628839
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOHRJ25TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .490
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHRJ25"
                     /clone_lib="BOHR"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       55 a    148 c    168 g    119 t
ORIGIN
```

```
  Query Match           11.4%;  Score 113.8;  DB 28;  Length 490;
  Best Local Similarity   54.1%;  Pred. No. 4.6e-16;
  Matches  259;  Conservative    0;  Mismatches  212;  Indels    8;  Gaps    1;

Qy        243 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 302
              |||||     || |||| | |||||     || |||| | |||||     || |||| |
Db          2 TCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGG 61

Qy        303 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 362
              |||||     | | |||| | |||||     || |||| | |||||     || |||| |
Db         62 TCGCTACGTTACATGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGG 121

Qy        363 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAG 422
              |||||     || |||| | |||||     || |||| | |||||     || |||| |
```

```
Db        122 TCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGG 181

Qy        423 GGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACG--------CCTGCCTGATGGC 474
              |  |    |  |||| | | |||| |  |   |  ||          || || | |
Db        182 TCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGG 241

Qy        475 TTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 534
              | | ||    |  |||| ||  ||||    |  |||| ||  ||||    |  |||| ||
Db        242 TCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGG 301

Qy        535 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGG 594
                ||||    |  |||| ||  ||||    |  | || ||  ||||    |  |||| ||
Db        302 TCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGG 361

Qy        595 CTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 654
              ||||    |  | || ||  ||||    |  |||| ||  ||||    |  |||| ||
Db        362 TCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGACGCTACGTAGCGTGCTCGG 421

Qy        655 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTG 713
                ||||    |  |||| ||  ||||       ||| |   |||   | || |||| |
Db        422 ACGCTACGTAGCGTGCTCGGACGCTACGTACCGTGCTCGGACGCTACGTAGCGTGCTCG 480
```

RESULT 40
BZ263814
LOCUS BZ263814 711 bp DNA linear GSS 15-OCT-2002
DEFINITION CH230-420O12.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone CH230-420O12, genomic survey sequence.
ACCESSION BZ263814
VERSION BZ263814.1 GI:23975613
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 711)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P. and Fraser,C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL Unpublished
COMMENT Other_GSSs: CH230-420O12.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230 (http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/or ering_information.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 420 row: O column: 12
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .711
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SsNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-420O12"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI; CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by Pieter de Jong"
BASE COUNT 149 a 157 c 191 g 214 t
ORIGIN

Query Match 11.3%; Score 113; DB 29; Length 711;
Best Local Similarity 71.3%; Pred. No. 7.8e-16;
Matches 149; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```
Qy        211 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 270
              || ||  | |  | ||| ||  |  ||   ||| ||| |  |||||||||||| ||| ||
Db          9 CAATAATGGTATTAGCTAGCCTGCTGGTGGGCTTGCTAGGTTGCTTGCTTGCTTGCTTGC 68

Qy        271 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 330
              |||||||||||| ||| || ||||||||||| |||  | ||||| ||||||  || ||
Db         69 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTCCTTGCTTCCTTGCTTCCTTGC 128

Qy        331 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 390
               | ||||||| || |||  | ||||| |||||| ||| ||||| | ||| ||| |||  |
Db        129 TTCCTTGCTTCCTTGCTTCCTTGCTTCCTTGCTTGCTTGCCTGATGGCTAGCTTGCTATC 188

Qy        391 CTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
               | ||| ||| ||  ||  | | ||| ||
Db        189 TTCCTTCCTTCCTTCCTTTCTTCCTTCCT 217
```

RESULT 41
BH725237
LOCUS BH725237 821 bp DNA linear GSS 20-FEB-2002
DEFINITION BOMHG42TF BO_2_3_KB Brassica oleracea genomic clone BOMHG42, genomic survey sequence.
ACCESSION BH725237
VERSION BH725237.1 GI:18830632
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 821)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished
COMMENT Other_GSSs: BOMHG42TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .821
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOMHG42"
/clone_lib="BO_2_3_KB"
/note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 86 a 284 c 280 g 171 t
ORIGIN

Query Match 11.3%; Score 113; DB 28; Length 821;
Best Local Similarity 49.3%; Pred. No. 8.1e-16;
Matches 329; Conservative 0; Mismatches 330; Indels 8; Gaps 1;

```
Qy         77 GCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTT 136
              ||  | |||||     |||||||  |  |  |    || |||| | |||||     |1 |
Db         14 GCTCGGTCGCTTCATAGCTTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGT 73

Qy        137 GCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 196
              |||  |  |  |    || |||| | |        ||   ||| | | |||   | || |
```

```
Db        74,GCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGT 133

Qy       197 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 256
             ||  | |       ||   | | | | |||   |||   | | | |||||     || |
Db       134 GCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGT 193

Qy       257 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 316
             ||  | |||||     || |||  | |||||     || |||  | ||| ||      || |
Db       194 GCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCTCTACGTAGCGT 253

Qy       317 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 376
             ||  | |||||     || |||  | |||||     || |||  | |||||     || |
Db       254 GCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGT 313

Qy       377 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 436
             ||  | |||||     || |||  | |||||     || |||   |  |  |  .|  |
Db       314 GCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGT 373

Qy       437 GCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGC--------TTGATAGTAGGGCT 488
             ||  | | |||| |  |   |  ||   ||    | ||        | | ||    |  |
Db       374 GCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGT 433

Qy       489 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 548
             ||  ||  ||||    |  |||  ||   |||    |  |||  ||  ||||    |  |
Db       434 GCCCGGTCGCTACGTAGCGTGCCCGGTCACTACGTAGCGTGCCCGGTCGCTACGTAGCGT 493

Qy       549 GCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 608
             ||  ||  ||||    |  | |  ||  ||||    |  |||  ||  ||||    |  |
Db       494 GCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGT 553

Qy       609 CCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 668
               |  ||  ||||    |  |||  ||  ||||    |  |||  ||  ||||    |  |
Db       554 GCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGT 613

Qy       669 GCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTT 728
             ||  ||  ||||    |  ||||  | |  |||   | ||  |||  | | |    |  || |
Db       614 GCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGT 673

Qy       729 GCCTGCT 735
             |||  | |
Db       674 GCCCGGT 680


RESULT 42
BH267859
LOCUS       BH267859                749 bp    DNA     linear   GSS 30-NOV-2001
DEFINITION  CH230-66L14.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
            CH230-66L14, genomic survey sequence.
ACCESSION   BH267859
VERSION     BH267859.1  GI:17180169
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 749)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 EcoRI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-66L14.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 66  row: L  column: 14
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .749
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-66L14"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 1"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      175 a    208 c    135 g    231 t
ORIGIN

  Query Match             11.3%;  Score 112.6;  DB 28;  Length 749;
  Best Local Similarity   72.9%;  Pred. No. 9.8e-16;
  Matches  145;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy       221 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 280
             ||| ||  |||| || ||  |||| | | || ||||||||||||| ||| || |||||||||
Db       482 GCTAGCCTGCTGGTGGGCTTGCTAGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 541

Qy       281 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 340
             |||  || || ||||| ||||||  || || | |||||| || |||  | ||||| |||
Db       542 GCTTTCTTGCTTGCTTACTTGCTTCCTTGCTTCCTTGCTTCCTTGCTTACTTGCTTCCTT 601

Qy       341 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
             |||  || ||.| ||||||| || ||| || ||| || | ||| ||| ||   ||| |||
Db       602 GCTTCCTTGCTTCCTTGCTTCCTTGCTTGCTTGCCTGATGGCTAGCTTGCTATCTTACTT 661

Qy       401 GCTCGCTGGCCTGCTTGCT 419
               || |||  | | ||| ||
Db       662 ACTTGCTTACTTTCTTCCT 680


RESULT 43
BH474846/c
LOCUS       BH474846                660 bp    DNA     linear   GSS 13-DEC-2001
DEFINITION  BOHLI64TF BOHL Brassica oleracea genomic clone BOHLI64, genomic
            survey sequence.
ACCESSION   BH474846
VERSION     BH474846.1  GI:17682957
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 660)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOHLI64TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .660
                     /organism="Brassica oleracea"
```

```
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHLI64"
                     /clone_lib="BOHL"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      112 a    217 c    250 g     81 t
ORIGIN

  Query Match             11.2%;  Score 111.6;  DB 28;  Length 660;
  Best Local Similarity   51.2%;  Pred. No. 1.6e-15;
  Matches  291;  Conservative    0;  Mismatches  269;  Indels    8;  Gaps    1;

Qy        155 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 214
              | || ||||  |  | ||    || |||| | ||||    | || ||||  |  | ||
Db        654 CAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACG 595

Qy        215 AGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 274
              |  |||| |. | || |    |  |||||| | |||     || ||||||  | |||
Db        594 CAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACG 535

Qy        275 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 334
              || |||||| | |||     || |||||| | |||     || |||||| | |||
Db        534 CAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACG 475

Qy        335 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 394
              || |||||| | |||     || |||||| | |||     || |||||| | |||
Db        474 CAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACG 415

Qy        395 TTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGA 454
              || |||||| | |||     ||       |    |    || || |||| | |
Db        414 CAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGT------- 362

Qy        455 CGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 514
              |      | || ||   | | |  |    |  |||| ||  || |    |  |||  ||
Db        361 -CGCCACGCAGCGTGCCCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCCCGG 303

Qy        515 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGG 574
              || |    |  |||| ||  || |    |  |||| ||  || |    |  | || ||
Db        302 TCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGG 243

Qy        575 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGG 634
              || |    |  |||| ||  || |    |  | || ||  || |    |  |||| ||
Db        242 TCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGG 183

Qy        635 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGC 694
              || |    |  |||| ||  || |    |  |||| ||  || |    |  ||| |
Db        182 TCGCCACGCAGCGTGCTCGGTCGCCATGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGG 123

Qy        695 TGGCTTGCTTGCTTGCTTGCTAGTGGGG 722
              | ||      || |||| | | || | |
Db        122 TCGCCACGCAGCGTGCTCGGTCGTAGCG 95


RESULT 44
BZ060388/c
LOCUS       BZ060388                732 bp    DNA     linear   GSS 10-OCT-2002
DEFINITION  l1g98b12.g1 B.oleracea002 Brassica oleracea genomic, genomic survey
            sequence.
ACCESSION   BZ060388
VERSION     BZ060388.1  GI:23669293
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 '(bases 1 to 732)
  AUTHORS   Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash
            ,W., Rabinowicz,P.D. and Wilson,R.K.
  TITLE     Whole genome shotgun reads from Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: l1g98  row: b  column: 12
            Seq primer: -21UPpOT forward
            Class: shotgun
            High quality sequence start: 8
            High quality sequence stop: 463.
FEATURES             Location/Qualifiers
     source          1. .732
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3712"
                     /clone_lib="B.oleracea002"
                     /note="Vector: pOTw13; Whole genome shotgun library from
                     flowering buds. DNA was purified  from a crude nuclear
                     prep using Brassica oleracea TO1000DH3 buds  provided by
                     Thosmas Osborn at the University of Wisconsin. Genomic
                     DNA was provided by Pablo Rabinowicz (CSHL) and the
                     shotgun library  prepared at Washington University Genome
                     Sequencing Center."
BASE COUNT      134 a    226 c    274 g     97 t      1 others
ORIGIN

  Query Match             11.2%;  Score 111.6;  DB 28;  Length 732;
  Best Local Similarity   51.2%;  Pred. No. 1.7e-15;
  Matches  291;  Conservative    0;  Mismatches  269;  Indels    8;  Gaps    1;

Qy        155 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 214
              | || ||||  |  | ||    || |||| | ||||    | || ||||  |  | ||
Db        723 CAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACG 664

Qy        215 AGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 274
              |  |||| | | || |    |  |||||| | |||     || ||||||  | |||
Db        663 CAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACG 604

Qy        275 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 334
              || |||||| | |||     || |||||| | |||     || |||||| | |||
Db        603 CAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACG 544

Qy        335 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 394
              || |||||| | |||     || |||||| | |||     || |||||| | |||
Db        543 CAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACG 484

Qy        395 TTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGA 454
              || |||||| | |||     ||       |    |    || || |||| | |
Db        483 CAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGT------- 431

Qy        455 CGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 514
              |      | || ||   | | |  |    |  |||| ||  || |    |  |||  ||
Db        430 -CGCCACGCAGCGTGCCCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCCCGG 372

Qy        515 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGG 574
              || |    |  |||| ||  || |    |  |||| ||  || |    |  | || ||
Db        371 TCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGG 312

Qy        575 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGG 634
              || |    |  |||| ||  || |    |  | || ||  || |    |  |||| ||
Db        311 TCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGG 252

Qy        635 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGC 694
              || |    |  |||| ||. || |    |  |||| ||  || |    |  ||| |
Db        251 TCGCCACGCAGCGTGCTCGGTCGCCATGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGG 192

Qy        695 TGGCTTGCTTGCTTGCTTGCTAGTGGGG 722
              | ||      || |||| | | || | |
Db        191 TCGCCACGCAGCGTGCTCGGTCGTAGCG 164


RESULT 45
AZ800556/c
LOCUS       AZ800556                465 bp    DNA     linear   GSS 16-FEB-2001
```

```
DEFINITION  2M0058F10R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0058F10 R, genomic survey sequence.
ACCESSION   AZ800556
VERSION     AZ800556.1  GI:12952820
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 465)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0058  row: F  column: 10
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 465.
FEATURES             Location/Qualifiers
     source          1. .465
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0058F10"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      172 a     98 c     95 g     99 t      1 others
ORIGIN

  Query Match             11.1%;  Score 110.8;  DB 28;  Length 465;
  Best Local Similarity   71.8%;  Pred. No. 2.3e-15;
  Matches  145;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        182 CTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTG 241
              | ||   |||  | |  ||  | |  ||| |       || ||| || ||  |  |  ||
Db        202 CATGGGAGCTCTCTTAATTAATCATTGGATAAATTCATTGATGGATTCCTTATTTGTTTG 143

Qy        242 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 301
              || ||| || ||||||||||||| ||| || ||| ||||||||| ||| || ||| ||||||
Db        142 CTTGCTTGCTTGCTTGCTTGCTTGCTTGCATGCTTGCTTGCTTGCATGCTTG 83

Qy        302 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 361
              |  ||| || ||||||||||||| ||| || ||||||||||  ||| || ||||||||||
Db         82 CATGCTTGCTTGCTTGCTTGCTTGCATGCTTGCTTGCATGCTTGCTTGCTTGCTTG 23

Qy        362 CTCGCTGGCCTGCTTGCTTGCT 383
              || ||| || ||| ||||||||
Db         22 CTTGCTTGCTTGCATGCTTGCT 1


RESULT 46
BZ470200
LOCUS       BZ470200               705 bp    DNA     linear   GSS 13-DEC-2002
DEFINITION  BONQQ85TF BO_1.6_2_KB_tot Brassica oleracea genomic clone BONQQ85,
            genomic survey sequence.
ACCESSION   BZ470200
VERSION     BZ470200.1  GI:26767003
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 705)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BONQQ85TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .705
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BONQQ85"
                     /clone_lib="BO_1.6_2_KB_tot"
                     /note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared
                     total DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      112 a     203 c     213 g     177 t
ORIGIN

  Query Match             11.0%;  Score 109.8;  DB 29;  Length 705;
  Best Local Similarity   54.3%;  Pred. No. 4.3e-15;
  Matches  222;  Conservative    0;  Mismatches  187;  Indels    0;  Gaps    0;

Qy         73 GCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 132
              || |||  | ||||||    || ||||  |  |       || |||| | ||||||
Db         41 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 100

Qy        133 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 192
              || ||||  |  |       || ||||  | |  .  |   ||   ||  | | |||   |
Db        101 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 160

Qy        193 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCT 252
              || |||| | |    |  ||   |  | | ||    |||   |   | ||||||
Db        161 GCGTGCTCGGTCGCAGCGTAGCGTGCTCGGTCGCAGCGTAGCGTGCTCGGTCGCTGCGTA 220

Qy        253 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 312
              || |||| | ||||||    || |||| | ||||||    || |||| | ||||||
Db        221 GCGTGCTCGGTCGCTGTGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 280

Qy        313 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 372
              || |||| | ||||||    || |||| | ||||||    || |||| | ||||||
Db        281 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 340
```

```
Qy        373 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAG 432
              || |||| | |||||     || |||| | ||||||    || ||||  |  |
Db        341 GCGTGCTCGGTCGCTCCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 400

Qy        433 GGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAG 481
              |  |||| | | |||  |  |   |  ||   ||    | || || | |
Db        401 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTCCGTAGCGTGCTCG 449


RESULT 47
BZ179889
LOCUS       BZ179889                416 bp    DNA     linear   GSS 11-OCT-2002
DEFINITION  CH230-444A16.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-444A16, genomic survey sequence.
ACCESSION   BZ179889
VERSION     BZ179889.1  GI:23824087
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-444A16.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 444  row: A  column: 16
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .416
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-444A16"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT       57 a    118 c     66 g    175 t
ORIGIN

  Query Match            10.9%;  Score 108.8;  DB 28;  Length 416;
  Best Local Similarity  64.1%;  Pred. No. 6.3e-15;
  Matches  164;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy        221 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 280
              ||| || |||| || ||  |||| | | || |||||||||||| ||| || |||||||||
Db         64 GCTAGCCTGCTGGTGGGCTTGCTAGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 123

Qy        281 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 340
              |||  ||  | ||||| ||||||  || || | ||||||| || |||  | ||||| |||
Db        124 GCTTTCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTT 183

Qy        341 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
              |||  || || ||||||| || | ||| || ||||  ||| ||  ||  | | ||| |||
Db        184 GCTTCCTTGCTTGCTTGCCTGATGGCTAGCTTGCTATCTTCCTTCCTTCCTTCCTTCCTT 243

Qy        401 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTAC 460
              ||  ||  | | ||| ||       |        ||||| ||| ||       |    |
Db        244 TCTTCCTTCCTTCCTTCCTTCCTTTCCTTCTTTATTGCTTTCTTTCTTTCTTTCCTTCTC 303

Qy        461 GCCTGCCTGATGGCTT 476
              | | | | ||  |||
Db        304 CCTTCCTTCATTCCTT 319


RESULT 48
AI676341
LOCUS       AI676341                438 bp    mRNA    linear   EST 19-MAY-1999
DEFINITION  etmEST0088 EtH1 Eimeria tenella cDNA clone etmb035 5', mRNA
            sequence.
ACCESSION   AI676341
VERSION     AI676341.1  GI:4876821
KEYWORDS    EST.
SOURCE      Eimeria tenella
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 438)
  AUTHORS   Wan,K.L., Chong,S.P., Ng,S.T., Tomley,F.M. and Jangi,M.S.
  TITLE     Survey of genes in Eimeria tenella merozoite
  JOURNAL   Unpublished
COMMENT     Contact: Wan KL
            Centre for Gene Analysis and Technology
            Universiti Kebangsaan Malaysia
            43600 UKM Bangi, Selangor DE, Malaysia
            Tel: 6 03 8292997
            Fax: 6 03 8293249
            Email: klwan@pkrisc.cc.ukm.my
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7
            Seq primer: SK.
FEATURES             Location/Qualifiers
     source          1. .438
                     /organism="Eimeria tenella"
                     /mol_type="mRNA"
                     /strain="Houghton"
                     /db_xref="taxon:5802"
                     /clone="etmb035"
                     /dev_stage="second generation merozoite"
                     /lab_host="XL1-Blue MRF'"
                     /clone_lib="EtH1"
                     /note="Vector: Lambda ZAPII; Site_1: EcoRI; Site_2: XhoI;
                     Second generation merozoites of E. tenella H were purified
                     by column  chromatography and mRNA extracted using a
                     FASTTRACK kit (Invitrogen).  cDNA was synthesised and a
                     Uni-ZAP XR library was constructed using  cDNA synthesis
                     kit, ZAP-cDNA synthesis kit and ZAP-cDNA Gigapack III
                     Gold cloning kit (Stratagene). The library was amplified
                     once through  E. coli XL1-Blue MRF'."
BASE COUNT       87 a    121 c    150 g     80 t
ORIGIN

  Query Match            10.9%;  Score 108.8;  DB 9;  Length 438;
  Best Local Similarity  64.9%;  Pred. No. 6.4e-15;
  Matches  161;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy        201 GCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTT 260
              ||||||| |       || | ||||||| |||||   |   ||| ||| || |||| | |
Db        150 GCTGAGGTGGTCTCGGGACCGCTGGCTAGCTAGCTAGCTAGCTAGCTTGCTGGATA 209

Qy        261 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 320
              ||| | | ||  ||| ||| ||| ||| ||  | | ||| ||| ||| ||  ||| |||
Db        210 GCTGGATAGCTGGCTAGCTGGCTGGCTAGCTGGGTCGCTGGCTAGCTAGCTGGCTAGCTC 269
```

```
Qy        321 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 380
              ||| ||| ||  ||| ||| ||| ||||||  ||| ||| ||| ||||||  ||| |||
Db        270 GCTGGCTAGCTAGCTGGCTAGCTAGCTGGCTAGCTAGCTGGCTAGCTGGCTGGCTAGCTA 329

Qy        381 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTT 440
              ||| ||| ||| |  ||| ||| ||||||  ||  | ||    | |     |   | |
Db        330 GCTAGCTAGCCACCAAGCTAGCTAGCTGGCTAGCGAGTTGCCTTGCCGCCCCGCTAGGTG 389

Qy        441 GCTTGCTA 448
              ||| ||||
Db        390 GCTAGCTA 397


RESULT 49
CNS04GRC
LOCUS       CNS04GRC                462 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            109C01 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL290001
VERSION     AL290001.1  GI:8028581
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 462)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .462
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="109C01"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG109AB01LP1-end : T7"
BASE COUNT       27 a      7 c    249 g    159 t     20 others
ORIGIN

  Query Match             10.9%;  Score 108.8;  DB 29;  Length 462;
  Best Local Similarity   52.5%;  Pred. No. 6.5e-15;
  Matches  208;  Conservative    9;  Mismatches  179;  Indels    0;  Gaps    0;

Qy        267 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 326
              | |  || |||  || | | | |  ||  || ||| | | : |  || :|| :|| : |
Db         58 TTGGATGGTTGGGTGGTTGGTTGGATGGNTGGTTGGTTGGBTGGKTGGKTGGBTGGKTGG 117

Qy        327 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 386
              | |  || |||  ||  | | |  ||  ||  || | | | |  || ||| ||| | |
Db        118 TTGGGTGGTTGGGTGGGTGGTTGGGTGGNTGGNTGGTTGGTTGGNTGGTTGGTTGGTTGG 177

Qy        387 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGC 446
              | |  || ||| ||| | | | |  || ||| || | ||   |  |||  | | | | |
Db        178 TTGNTTGGTTGGTTGGTTGGTTGGTTGGTTGNTGGGTGGGTGGGTGGGTGGGTGGGTGGG 237

Qy        447 TAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGG 506
              |  | ||  ||.  ||  |   | ||| | | | |  |||  | |||| :| | |   ||
Db        238 TGGGTGGGTGGGTGGCTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTKGGTGGGTTGG 297

Qy        507 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 566
               || |||   | | || ||   | |||:  | | |: ||:  | |||   | | || ||
Db        298 TTGGTGGTTGGTTKGTTGGTTGGTTGGKTGGKTGGKTGGKTGGTTGGTTGGKTGGTTGGT 357

Qy        567 CTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGG 626
                |||| | | |  ||| | |||| || | |  ||    |||| || | |  ||
Db        358 TGGTTGGGTGGTTGGGTGGGTGGGTGGGTTGGTGGGTGGTTGGGTGGGTTGTTGGCTGGC 417

Qy        627 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGT 662
              |:|||   ||| || |||  |||||   ||| ||
Db        418 TGGYTGGCTGGCTGGTTGGGTGGCTGGTTGGCTGGT 453


RESULT 50
BH700477
LOCUS       BH700477                528 bp    DNA     linear   GSS 20-FEB-2002
DEFINITION  BOMLB36TF BO_2_3_KB Brassica oleracea genomic clone BOMLB36,
            genomic survey sequence.
ACCESSION   BH700477
VERSION     BH700477.1  GI:18776991
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOMLB36TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOMLB36"
                     /clone_lib="BO_2_3_KB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       81 a    138 c    178 g    131 t
ORIGIN

  Query Match             10.9%;  Score 108.8;  DB 28;  Length 528;
  Best Local Similarity   56.6%;  Pred. No. 6.7e-15;
  Matches  223;  Conservative    0;  Mismatches  167;  Indels    4;  Gaps    1;

Qy        205 AGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTC 264
```

```
Db        134 AGCGCAAAAGCTGGTCACTACGTAGCGTGCTCGGTCGCTACGTGGCGTGCTCTGTCGCTA 193

Qy        265 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 324

Db        194 CGTGGCGTGCTCGGTCGCTACGTGGCGTGCTCGGTCGCTACATGGCGTGCTCGGTCGCTA 253

Qy        325 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 384

Db        254 CGTGGCTTGCTCGGTCGCTACGTGGCGTGCTCGGTCGCTACGTGGCGTGCTCGGTCGCTC 313

Qy        385 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGC----TTGCTGAGGGGACAGTAGGGCTGCTT 440

Db        314 CGTGGCGTGCTCGGTCGCTACGTGGCGTGCTCGGTCGCTACGTGGCGTGCTCGGTCGCTA 373

Qy        441 GCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTA 500

Db        374 CGTGGCGTGCTCGGTCGCTACGTGGCGTGTTCGGTCGCTACGTTGCGTGTTCGGTCGCTA 433

Qy        501 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 560

Db        434 CGTGGCGTGCTCGGTCGCTACGTGGCGTGCTCGGTCGCTACGTGGCGTGCTCGGTCGCTA 493

Qy        561 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGG 594

Db        494 CGTGGCGTGCTCGGTCGCTACGTGGCGTGCTGGG 527
```

```
RESULT 51
BH544637
LOCUS       BH544637                794 bp    DNA     linear   GSS 14-DEC-2001
DEFINITION  BOGGJ62TF BOGG Brassica oleracea genomic clone BOGGJ62, genomic
            survey sequence.
ACCESSION   BH544637
VERSION     BH544637.1  GI:17796418
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 794)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOGGJ62TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .794
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGGJ62"
                     /clone_lib="BOGG"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      157 a    235 c    235 g    167 t
ORIGIN

  Query Match          10.8%;  Score 108.4;  DB 28;  Length 794;
  Best Local Similarity   53.4%;  Pred. No. 9.3e-15;
  Matches  251;  Conservative    0;  Mismatches  216;  Indels    3;  Gaps    1;

Qy         70 GCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGG 129
```

```
Db        328 GCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTAC 387

Qy        130 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCG 189

Db        388 GCAGCGTGCTTGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTAC 447

Qy        190 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGG 249

Db        448 GCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCCCGGTCGCTAC 507

Qy        250 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 309

Db        508 GCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTAC 567

Qy        310 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 369

Db        568 GCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTAC 627

Qy        370 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 429

Db        628 GCAGCGTGCCCGGTCGCTATGCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTAC 687

Qy        430 TAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTG 489

Db        688 GCAGCATGCCCGGTCGCTACGCAGCATG---CTCAGCGTGCTCGGTCGCTACGCAGCGTG 744

Qy        490 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 539

Db        745 CTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCT 794
```

```
RESULT 52
BH483727
LOCUS       BH483727                569 bp    DNA     linear   GSS 13-DEC-2001
DEFINITION  BOGCW50TF BOGC Brassica oleracea genomic clone BOGCW50, genomic
            survey sequence.
ACCESSION   BH483727
VERSION     BH483727.1  GI:17691831
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 569)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOGCW50TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .569
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGCW50"
                     /clone_lib="BOGC"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       69 a    178 c    188 g    134 t
ORIGIN

  Query Match          10.8%;  Score 107.8;  DB 28;  Length 569;
```

```
  Best Local Similarity   52.5%;  Pred. No. 1.2e-14;
  Matches  265;  Conservative    0;  Mismatches  232;  Indels    8;  Gaps    1;

Qy        221 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 280
              | | |||   |||    |   | |||||     || |||| | |||||      || ||||
Db         24 GGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTC 83

Qy        281 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 340
              |  |||       || |||| | || ||      || |||| | || ||      || ||||
Db         84 GGCCGCAATGTAGCGTGCTCGGTCCCTACGTAGCGTGCTCGGTCCCTACGTAGCGTGCTC 143

Qy        341 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
              | || ||      || |||| | || ||      || |||| | |||||      || ||||
Db        144 GGTCCCTACGTAGCGTGCTCGGTCCCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTC 203

Qy        401 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTAC 460
              |  || |      || ||||  |  || |     |  |||| | | |||| |  |   |  |
Db        204 GGACGATACGTAGCGTGCTCGGACGATACGTAGCGTGCTCGGTCGCTACGTAGCATGCTC 263

Qy        461 G--------CCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 512
              |        || || | | | | ||    |  |||| ||  ||||    |  ||||
Db        264 GGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTC 323

Qy        513 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTG 572
              ||  ||||    | ||||  ||  ||||    |  ||||  ||  ||||    |  | ||
Db        324 GGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTC 383

Qy        573 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTG 632
              ||  ||||    | ||||  ||  ||||    |  | ||  ||  ||||    |  ||||
Db        384 GGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTC 443

Qy        633 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCT 692
              ||  ||||    | ||||  ||  ||||    |  ||||  ||  ||||    |  ||
Db        444 GGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGGTC 503

Qy        693 GCTGGCTTGCTTGCTTGCTTGCTAG 717
              | | |||   | || |||| | | |
Db        504 GGTCGCTACGTAGCGTGCTCGGTCG 528


RESULT 53
CB952830
LOCUS       CB952830                819 bp    mRNA    linear   EST 29-APR-2003
DEFINITION  AGENCOURT_13688077 NIH_MGC_176 Mus musculus cDNA clone
            IMAGE:30304742 5', mRNA sequence.
ACCESSION   CB952830
VERSION     CB952830.1  GI:30208948
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 819)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein
             cDNA Library Preparation: Michael Brownstein Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM77  row: f  column: 15
            High quality sequence stop: 212.
FEATURES             Location/Qualifiers
     source          1. .819
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30304742"
                     /lab_host="DH10B (T1-phage-resistant)"
                     /clone_lib="NIH_MGC_176"
                     /note="Organ: kidney; Vector: pDNR-LIB; Site_1: SfiI
                     (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc); cDNA made
                     by oligo-dT priming and directionally cloned. 5' and 3'
                     adaptors were used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Full-length
                     enriched library was constructed using the Clontech
                     Creator SMART kit and size-selected to contain the 0.5 kb
                     size fraction. Library created in the laboratory of M.
                     Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."
BASE COUNT       93 a    206 c    242 g    277 t      1 others
ORIGIN

  Query Match           10.8%;  Score 107.8;  DB 14;  Length 819;
  Best Local Similarity   57.6%;  Pred. No. 1.3e-14;
  Matches  193;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy        223 TGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 282
              ||| | | | || ||   | | |||||  || | ||| ||| | |||   ||| | | |
Db         46 TGGTTGGTTGGTTGGTTGGTTGGCTGGTTTGTTGGCTGGCTGGTTGGTTGGCTGGTTGGT 105

Qy        283 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 342
              | |||||   ||| ||| ||| | ||||  | | | | | | | |||   | | | | |
Db        106 TGGCTGGTTGGCTGGCTGGCTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGT 165

Qy        343 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 402
              | | |||   ||| ||| ||| |||||   ||| | | | | | |||   ||| | | |
Db        166 TGGTTGGTTGGCTGGCTGGCTGGCTGGTTGGCTGGTTGGTTGGTTGGTTGGCTGGTTGGT 225

Qy        403 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGC 462
              | ||||||  ||| | ||   ||  || ||   | | | | |   ||  ||   |
Db        226 TGGCTGGCTGGCTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGGA 285

Qy        463 CTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGT 522
                ||  | ||| | ||| || ||   |||||   |    | || | |||| || ||| ||
Db        286 TGGCTGGCTGGTTGGATGGTTGGTTGGCTGGCTGGTCGATTGGCCGGCTGCGCGGCTGGT 345

Qy        523 AGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTG 557
              |||||  | |||||   ||  || ||   || ||
Db        346 TTGGCTGGTTGGCTGGCTGTTTGGTTGGCTGGTTG 380


RESULT 54
BH677755
LOCUS       BH677755                677 bp    DNA     linear   GSS 19-FEB-2002
DEFINITION  BOHYJ15TR BO_2_3_KB Brassica oleracea genomic clone BOHYJ15,
            genomic survey sequence.
ACCESSION   BH677755
VERSION     BH677755.1  GI:18748198
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 677)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .677
                     /organism="Brassica oleracea"
```

```
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHYJ15"
                     /clone_lib="BO_2_3_KB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      101 a    202 c    222 g    152 t
ORIGIN

  Query Match            10.7%;  Score 106.6;  DB 28;  Length 677;
  Best Local Similarity  53.3%;  Pred. No. 2.3e-14;
  Matches  295;  Conservative    0;  Mismatches  249;  Indels    9;  Gaps    3;

Qy        177 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 236
              || |||| | |||||     || ||||  | |   |    |  |||| |   ||||
Db        125 GCATGCTCGGTCGCTACGTAGCGTGCTCGGCGACTACGTAGCGTGCTCGGCCGCTACGTA 184

Qy        237 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 296
                ||||||  | ||    | || |||||| | ||    | || |||||| | ||    |
Db        185 CCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 244

Qy        297 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 356
              || ||||||   ||    | || |||||| |  |    | || |||||| |  |    |
Db        245 GCGTGCTCGGCCGCAATGTAGCGTGCTCGGTCCCTACGTAGCGTGCTCGGTCCCTACGTA 304

Qy        357 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 416
              || ||||||  |  |    | || |||||| |  |    | || |||||| | ||    |
Db        305 GCGTGCTCGGTCCCTACGTAGCGTGCTCGGTCCCTACGTAGCGTGCTCGGTCGCTACGTA 364

Qy        417 GCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTT 476
              ||      |  |    |  || ||||        ||||| ||||   || || | | |
Db        365 GCGTGCTCGGACGATACGTAGCGTGCT-------CGGACGATACG-TAGCGTGCTCGGTC 416

Qy        477 GATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 536
              | ||    |  |||| ||  ||||    |  |||| ||  ||||    |  |||| ||
Db        417 GCTACGTAGCATGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTC 476

Qy        537 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCT 596
              ||||    |  |||| ||  ||||    |  | || ||  ||||    |  |||| ||
Db        477 GCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTC 536

Qy        597 GCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 656
              ||||    |  | || ||  ||||     |||||| ||  ||||    |  |||| ||
Db        537 GCTACGTAGCGTGCTCGGTCGCTA-CGTAGCTGCTCGGTCGCTACGTAGCGTGCTCGGTC 595

Qy        657 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTA 716
              ||||    |  |||| ||  ||||    |  |||  | | |||   | || |||| | |
Db        596 GCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTC 655

Qy        717 GTGGGGCCGCTTG 729
              |    |  || ||
Db        656 GCTACGTAGCGTG 668


RESULT 55
BZ497497
LOCUS       BZ497497                748 bp    DNA     linear   GSS 16-DEC-2002
DEFINITION  BONKB78TR BO_1.6_2_KB_tot Brassica oleracea genomic clone BONKB78,
            genomic survey sequence.
ACCESSION   BZ497497
VERSION     BZ497497.1  GI:27010919
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 748)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BONKB78TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .748
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BONKB78"
                     /clone_lib="BO_1.6_2_KB_tot"
                     /note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared
                     total DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      153 a    210 c    218 g    167 t
ORIGIN

  Query Match            10.7%;  Score 106.6;  DB 29;  Length 748;
  Best Local Similarity  55.1%;  Pred. No. 2.4e-14;
  Matches  231;  Conservative    0;  Mismatches  184;  Indels    4;  Gaps    1;

Qy        161 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 220
              |||  |  |  |     || |||| | |||||      || ||||  |  | ||    | |
Db        317 GCTCGGTAGCTACGTAGCATGCTCGGTCGCTACGTAGCGTGCTCGGTCGCCACGTAGCGT 376

Qy        221 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 280
              ||| | | || |     |  |||||| | |||    | || ||||||  | |||    | || |
Db        377 GCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGT 436

Qy        281 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 340
              |||||  | |||    | || ||||||  | |||    | || ||||||  | |||    | || |
Db        437 GCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGT 496

Qy        341 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
              |||||  | |||    | || ||||||  | |||    | || ||||||  | |||    | || |
Db        497 GCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGT 556

Qy        401 GCTCGCTGGCC----TGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACG 456
              |||||  | |||     || ||||  |  | ||    |  |||| | | || | |  |
Db        557 GCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGT 616

Qy        457 GTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 516
              |  ||   |||   | || || | |   | |   | |  |||| |   | |   | |  |
Db        617 GCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGT 676

Qy        517 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGC 575
              ||| |   | |   | |  |||| |   | |   | |  || | |   | | | |||||
Db        677 GCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGATCGGTCGCCACGTGGGC 735


RESULT 56
AI676338
LOCUS       AI676338                457 bp    mRNA    linear   EST 19-MAY-1999
DEFINITION  etmEST0084 EtH1 Eimeria tenella cDNA clone etmb031 5', mRNA
            sequence.
ACCESSION   AI676338
VERSION     AI676338.1  GI:4876818
KEYWORDS    EST.
SOURCE      Eimeria tenella
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   Wan,K.L., Chong,S.P., Ng,S.T., Tomley,F.M. and Jangi,M.S.
  TITLE     Survey of genes in Eimeria tenella merozoite
  JOURNAL   Unpublished
COMMENT     Contact: Wan KL
            Centre for Gene Analysis and Technology
```

Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor DE, Malaysia
Tel: 6 03 8292997
Fax: 6 03 8293249
Email: klwan@pkrisc.cc.ukm.my
PCR PRimers
FORWARD: T3
BACKWARD: T7
Seq primer: SK.

FEATURES Location/Qualifiers
source 1. .457
/organism="Eimeria tenella"
/mol_type="mRNA"
/strain="Houghton"
/db_xref="taxon:5802"
/clone="etmb031"
/dev_stage="second generation merozoite"
/lab_host="XL1-Blue MRF'"
/clone_lib="EtH1"
/note="Vector: Lambda ZAPII; Site_1: EcoRI; Site_2: XhoI; Second generation merozoites of E. tenella H were purified by column chromatography and mRNA extracted using a FASTTRACK kit (Invitrogen). cDNA was synthesised and a Uni-ZAP XR library was constructed using cDNA synthesis kit, ZAP-cDNA synthesis kit and ZAP-cDNA Gigapack III Gold cloning kit (Stratagene). The library was amplified once through E. coli XL1-Blue MRF'."

BASE COUNT 93 a 122 c 155 g 87 t
ORIGIN

Query Match 10.6%; Score 106.4; DB 9; Length 457;
Best Local Similarity 67.7%; Pred. No. 2.3e-14;
Matches 149; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy      201 GCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTT 260
            ||||||| |      || | ||||||| |||||   |   ||| ||| || |||| | |
Db      146 GCTGAGGTGGTCTCGGGACCGCTGGCTAGCTAGCTAGCTAGCTAGCTTGCTGGATA 205

Qy      261 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 320
            ||| | | ||  ||| ||| ||| ||| ||  | | ||| ||| ||| ||  ||| |||
Db      206 GCTGGATAGCTGGCTAGCTGGCTGGCTAGCTGGGTCGCTGGCTAGCTAGCTGGCTAGCTC 265

Qy      321 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 380
            ||| ||| ||  ||| ||| ||| ||||||  ||| ||| ||| ||||||  ||| |||
Db      266 GCTGGCTAGCTAGCTGGCTAGCTAGCTGGCTAGCTAGCTGGCTAGCTGGCTGGCTAGCTA 325

Qy      381 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 420
            ||| ||| |||  |  ||| ||| ||||||  ||  | ||
Db      326 GCTAGCTAGCCACCAAGCTAGCTAGCTGGCTAGCGAGTTG 365
```

RESULT 57
CD347163/c
LOCUS CD347163 566 bp mRNA linear EST 29-MAY-2003
DEFINITION EtESTef11c03.y1 Eimeria tenella M5-6 Excised cDNA Eimeria tenella cDNA 5' similar to TR:Q25062 Q25062 PRECURSOR PROTEIN FOR RFAMIDE. ;, mRNA sequence.
ACCESSION CD347163
VERSION CD347163.1 GI:31138784
KEYWORDS EST.
SOURCE Eimeria tenella
ORGANISM Eimeria tenella
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae; Eimeria.
REFERENCE 1 (bases 1 to 566)
AUTHORS Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T., Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
TITLE WashU-Merck Eimeria tenella project
JOURNAL Unpublished
COMMENT Contact: David Sibley, Ph.D.

WashU-Merck Eimeria tenella project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Contact David Sibley (toxoest@borcim.wustl.edu) for further information relating to organism, libraries, or clone availability.
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco.

FEATURES Location/Qualifiers
source 1. .566
/organism="Eimeria tenella"
/mol_type="mRNA"
/db_xref="taxon:5802"
/dev_stage="LS18"
/lab_host="E.coli DH10B (GeneHog, Invitrogen,Inc)"
/clone_lib="Eimeria tenella M5-6 Excised cDNA"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Sporozoites were obtained from in vitro sporulated and excysted oocysts of E.tenella grown in chickens. cDNA was synthesized from poly mRNA using an oligo-dT primer containing a XhoI site. Following second strand synthesis, EcoRI adapters were ligated to the cDNA and products were size-selected on sephacryl S500. The cDNA were ligated to EcoRI/XhoI prepared lambda ZapII (Stratagene). The primary library was mass excised using ExAssist helper phage (Stratagene). The phagemids were preciptated with PEG 8000 , extracted with phenol/chloroform and electroporated into DH10B cells. The library may contain a small percentage of host or bacterial contaminants. Library materials provided by: Paul Liberator, Merck Research Labs Library constructed by: Paul Liberator, Merck Research Labs."

BASE COUNT 174 a 186 c 134 g 72 t
ORIGIN

Query Match 10.4%; Score 104.4; DB 14; Length 566;
Best Local Similarity 51.5%; Pred. No. 7.1e-14;
Matches 240; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

```
Qy      113 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACA 172
            ||  | | | ||||| |   ||| ||||   ||   |   |   ||| ||||   ||
Db      564 GCTAGTTGGTTCGCTAGATGGCTCGCTGGTTGGTACGCTAGATGGCTCGCTGGTTGGTTC 505

Qy      173 GTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTA 232
            |   |   ||| | | | ||    ||| | ||    || ||| ||   ||  ||| |||
Db      504 GCCAGGTGGCTCGATAGTTGTTTCGCTAGATGGCTCGATAGTTGGTTCGCCAGCTGGCTT 445

Qy      233 GTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 292
            |  |||    | ||| |     | |||  || | | ||  | |   | ||| | |||
Db      444 GCTGGGTGTTTGGCTAGGTGAATCGCTAACTGGTTCGCTGGGTGATTCGCTAGTTGGTTC 385

Qy      293 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 352
            | | | | ||||||| |   | | |||   | | | ||  | | | | ||| | |||
Db      384 GATAGGTGGCTCGCTAGTTGGTTCGCTAATTGGTTCGCTAGTTGGTTCGCTAGTTGGTTC 325

Qy      353 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 412
            ||| ||| |||||||||  || | ||| | | | |||   |||  || ||| | ||
Db      324 GCTAGCTGGCTCGCTGGGTTGTTCGCTAGTTGGGTGGGTAGCTGACTCGCTGGGTGTTTG 265

Qy      413 GCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATG 472
            ||| | |||| | || ||    || || ||      | |||| ||| ||| | |
Db      264 GCTAGGTGAGTCGCTAGCTGGCTCTCTATCTGTCTCGCTAGCCGGTCCGCTAGCCGGCTC 205

Qy      473 GCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 532
            ||| |    |    |||    |||||  ||   ||  | ||    | | ||
Db      204 GCTGGTGTAAAAACTGTCTGACTCGCTAGCTGGTTCGCGTGACTCACTGAACGGTACGCT 145

Qy      533 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGC 578
             || |||     ||||  |  |||    |    |||    |||| |
Db      144 AGCGGCTTCACTGGCTAGTTCGCTAGCCGGCTCGCTAGCTGGCTTC 99
```

```
RESULT 58
CB504454/c
LOCUS       CB504454                710 bp    mRNA    linear   EST 16-MAY-2003
DEFINITION  ssalplnb502144_rev gut Salmo salar cDNA, mRNA sequence.
ACCESSION   CB504454
VERSION     CB504454.1  GI:29315680
KEYWORDS    EST.
SOURCE      Salmo salar (Atlantic salmon)
  ORGANISM  Salmo salar
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
REFERENCE   1  (bases 1 to 710)
  AUTHORS   GRASP Consortium, Davidson,W.S., Koop,B.F. and
            http://web.uvic.ca/cbr/grasp.
  TITLE     A survey of Salmo salar transcripts from high complexity cDNA
            libraries
  JOURNAL   Unpublished
COMMENT     Contact: Koop BF
            Centre for Biomedical Research
            University of Victoria
            PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
            Tel: 250 472 4067
            Fax: 250 472 4075
            Email: bkoop@uvic.ca
            Genome Sciences Centre, BC Cancer Agency cDNA preparation,
            sequencing and bioinformatics: Y Butterfield, R Kirkpatrick, J
            Asano, N Girn, R Guin, D Lee, S Lee, T Olson, P Pandoh, A Prahbu, D
            Smailus, L Spence, J Stott, S Taylor, G Yang, J Schein, S Jones and
            M Marra.
FEATURES             Location/Qualifiers
     source          1. .710
                     /organism="Salmo salar"
                     /mol_type="mRNA"
                     /strain="McConnell"
                     /db_xref="taxon:8030"
                     /clone_lib="gut"
                     /note="Vector: pBlueScriptIISK+; Library Creator:  Matthew
                     L Rise ; Atlantic salmon tissue contributors: Carlo Biagi,
                     Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.), Simon
                     Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery (Crofton,
                     B.C.), Rachel Roper (University of Victoria)"
BASE COUNT      172 a    210 c    184 g    144 t
ORIGIN

  Query Match            10.4%;  Score 104.2;  DB 14;  Length 710;
  Best Local Similarity  61.0%;  Pred. No. 8.4e-14;
  Matches  169;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy        157 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAG 216
              ||| |||||  ||   |  |||  ||  ||| ||||||  ||| |||||       | |
Db        682 GCTGGCTGACTGGCTGGCTGGCTGGCAGGCTGGCTGGCTGGCTGGCTGACTAACTGGCTG 623

Qy        217 GGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 276
              |   ||||||| ||| |  ||   ||| ||||||   ||  ||  ||  ||| |  |||
Db        622 GCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGACTAACTGACTGACTGACTGGCTG 563

Qy        277 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 336
               ||  || ||||||  |||  || |||  |||||  |||  || |||  ||  |   ||
Db        562 ACTAACTGGCTGGCTGGCTGACTGGCTGACTGGCTGGCTGACTGGCTGACTAACTGACTG 503

Qy        337 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 396
               || ||| |||| |  |||  ||  ||  ||| |  ||| ||| ||| |||| |   ||
Db        502 ACTGGCTGGCTGACTGGCTGACTAACTGACTGACTGGCTGGCTGGCTGGCTGACTAACTG 443

Qy        397 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGG 433
              ||| ||| ||||||  ||| |||||  ||   |  ||
Db        442 GCTGGCTGGCTGGCTGGCTGGCTGACTGGCTGGCTGG 406


RESULT 59
AK080786
LOCUS       AK080786                4346 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
            enriched library, clone:B130055L14 product:WASP family 1, full
            insert sequence.
ACCESSION   AK080786
VERSION     AK080786.1  GI:26099478
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 4346)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
```

```
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. Tomohiro Kono (Department of Animal
            Science, Tokyo University of Agriculture, 1737 Hunako Atsugi City,
            Kanagawa Prefecture, Japan) whose assistance we gratefully
            acknowledge.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .4346
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:B130055L14"
                     /db_xref="taxon:10090"
                     /clone="B130055L14"
                     /tissue_type="parthenogenote"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="9.5 days embryo"
     misc_feature    1. .4346
                     /note="WASP family 1 (MGD|MGI:1890563, GB|NM_031877,
                     evidence: BLASTN, 99%, match=394)"
BASE COUNT     1075 a    854 c   1012 g   1405 t
ORIGIN

  Query Match             10.3%;  Score 102.6;  DB 11;  Length 4346;
  Best Local Similarity   57.2%;  Pred. No. 3.1e-13;
  Matches  226;  Conservative    0;  Mismatches  164;  Indels    5;  Gaps    2;

Qy        174 TAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAG 233
              | | |  | |   | |||| || ||| ||||   ||   || ||   | ||| | | | |
Db        866 TCGACTGGTTGATTGGCTGCCCGGCTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGG 925

Qy        234 TAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 293
              | ||   | | | || |  ||| |||  ||  ||||   ||| ||| ||  | |   | |
Db        926 TTGGTTGGTTGGTTGTCTGGCTGGCTACCTGCCTGGTTGGCTGGCTAGCCTGGTT--CGG 983

Qy        294 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 353
                ||| | | | ||| | | | ||| ||| ||| ||  ||| ||| ||| |||| |  |
Db        984 TCGGCTGGGTGGTTGGTCGGTTGGCTGGCTGGCTAGCTGGCTGGCTGGCTGACTGG 1043

Qy        354 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 413
              || ||| ||| |||||| ||| ||| | | | |||   | | | | ||| |||||  |
Db       1044 CTGGCTGGCTGGCTGGCTGGCTGGTTGGATGGTTGGTTGGGTGGCTGGCTGGCTGG 1103

Qy        414 CTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGG 473
              || ||||  || | |  || ||  ||||| |  || | | |   | ||| ||| | |
Db       1104 CTGGCTGGCTGGGCTGGTTGGTTG---GCTTGGTTGGGTGGCTGGCTGGCTGGCTGTTTG 1160

Qy        474 CTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 533
                |||  |  |||||  |||||  |  ||  ||||  ||| |  |    ||
Db       1161 GTTGGGTGACCGGCTGGCTGGCTGACTGATTGGTGTCTGGCAGGCTGGCTAGTTGTGTGC 1220

Qy        534 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 568
```

```
                ||| |  |  | |||||  ||| || ||| |
Db       1221 TTGGCTGGCTGACCAGCTGGCTGGCTGGTTGGGTT 1255


RESULT 60
BX173189/c
LOCUS       BX173189               551 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-174I13, genomic survey sequence.
ACCESSION   BX173189
VERSION     BX173189.1  GI:28004894
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 551)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 174I13. 174I13
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .551
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-174I13"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
BASE COUNT      230 a    162 c     51 g    108 t
ORIGIN

  Query Match             10.2%;  Score 102.4;  DB 29;  Length 551;
  Best Local Similarity   52.3%;  Pred. No. 2e-13;
  Matches  226;  Conservative    0;  Mismatches  206;  Indels    0;  Gaps    0;

Qy        237 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 296
              | |||   |   | |||  ||  ||   |   | ||| ||| ||| | | | |  || ||
Db        539 GACTGACTGACTGACTGACTGACTGACTGATTGATTGATTGATTGATTGATT 480

Qy        297 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 356
              | ||| | | | |  || ||| ||| | | | |  || ||| ||| | | | |  || ||
Db        479 GATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATT 420

Qy        357 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 416
              | ||| | | | |  || ||| ||| | | | |  || ||| ||| | | | |  || ||
Db        419 GATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATT 360

Qy        417 GCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTT 476
              |    | | | |   ||||| ||| ||| | |  |       | |   |||| |  |
Db        359 GGCTGGCTGGCTGGCTGGCTGGTTGGTTGGTTGGTTGGTTAACTGGTTCGTTGATTGAAT 300

Qy        477 GATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 536
              | | |   |  || | || || |     || || | || || | |   || || | | |
Db        299 GGTTGGTTGCATGGTTGGATGGTTAGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 240

Qy        537 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCT 596
              | | |   || || | || || | |   || |  | || || | |   || || |   |
Db        239 GGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTAATT 180

Qy        597 GCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 656
              | | |   | ||   | |||   |   | |||   | |||   |   | |||   | ||
Db        179 GGTTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACT 120

Qy        657 GCTAGTAGGGCT 668
              | | |  || ||
Db        119 GATGGACGGACT 108
```

```
RESULT 61
BZ457666/c
LOCUS       BZ457666               787 bp    DNA     linear   GSS 13-DEC-2002
DEFINITION  BONRC33TF BO_1.6_2_KB_tot Brassica oleracea genomic clone BONRC33,
            genomic survey sequence.
ACCESSION   BZ457666
VERSION     BZ457666.1  GI:26737270
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 787)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BONRC33TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .787
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BONRC33"
                     /clone_lib="BO_1.6_2_KB_tot"
                     /note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared
                     total DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      169 a    246 c    238 g    134 t
ORIGIN

  Query Match           10.2%;  Score 102.4;  DB 29;  Length 787;
  Best Local Similarity   50.2%;  Pred. No. 2.2e-13;
  Matches  282;  Conservative    0;  Mismatches  276;  Indels    4;  Gaps    1;

Qy        178 CCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGG 237
              |||| | |||   | ||      |||      | ||   || |||   | ||| |   |
Db        616 CCTGGTCGCTACGTAGCGACTGAGCTCTCACCAAAGCTCGGTCGCTTCATAGCTTGCTCG 557

Qy        238 GCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 297
              |  |||   | || |||| | | |||   | || |||| | | |||   | || |||| |
Db        556 GTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCG 497

Qy        298 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 357
              | |||   | || |||  | | |||   | || |||  | | |||   | || |||  |
Db        496 GTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCG 437

Qy        358 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 417
              | |||   | || |||  | | |||   | || |||  | | |||   | || |||  |
Db        436 GTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCG 377

Qy        418 CTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCC----TGCCTGATGG 473
              |        ||  | | | |  ||   ||  | | | |  |||      || ||   |
Db        376 GTCGCTACGTAGCGTGCCCGGTCTCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCG 317

Qy        474 CTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 533
              | | ||    |  |||  ||  ||||    |  |||  ||  ||||    |  |||  |
Db        316 GTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCG 257

Qy        534 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGG 593
              |  ||||    |  |||  ||  ||||    |  | |  ||  ||||    |  |||  |
Db        256 GTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCG 197

Qy        594 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 653
              |  ||||    |  | |  ||  ||||    |  |||  ||  ||||    |  |||  |
Db        196 GTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCG 137

Qy        654 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTG 713
              |   |||    |  |||  ||  ||||    |  |||| | | |||   | || |||  |
Db        136 GTCACTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCG 77

Qy        714 CTAGTGGGGCCGCTTGCCTGCT 735
              | |    |  || |||| | |
Db         76 GTCGCTACGTAGCGTGCCCGGT 55


RESULT 62
CNS02WKU
LOCUS       CNS02WKU               881 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            176H16 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL217191
VERSION     AL217191.1  GI:7876010
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 881)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .881
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="176H16"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG176DD08SP1-end :
                     PUC-Ori"
BASE COUNT      104 a    217 c    166 g    311 t     83 others
ORIGIN

' Query Match           10.2%;  Score 101.6;  DB 29;  Length 881;
  Best Local Similarity   55.6%;  Pred. No. 3.5e-13;
  Matches  189;  Conservative    3;  Mismatches  148;  Indels    0;  Gaps    0;

Qy        250 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 309
```

```
Db        542 CCTATTTTTHTRCCTGCYTGCCTGCCTGTCTGTCTGCCTGACTGCCTGCCTGCCTGCCTG 601

Qy        310 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 369

Db        602 TCTGTCTGCCTGACTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTGTCTG 661

Qy        370 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 429

Db        662 CCTGACTGCCTGTCTGTCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTG 721

Qy        430 TAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTG 489

Db        722 TCTGCCTGACTGCCTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTGTCTG 781

Qy        490 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 549

Db        782 CCTGACTGCCTGTCTGTCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTGTCTGCCTG 841

Qy        550 CTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTG 589

Db        842 ACTGCCTGTCTGTCTGCCTGACTGCCTGTCTGTCTGCCTG 881


RESULT 63
BH423105/c
LOCUS       BH423105                622 bp    DNA     linear   GSS 12-DEC-2001
DEFINITION  BOHJE88TF BOHJ Brassica oleracea genomic clone BOHJE88, genomic
            survey sequence.
ACCESSION   BH423105
VERSION     BH423105.1  GI:17608833
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 622)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOHJE88TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .622
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHJE88"
                     /clone_lib="BOHJ"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      144 a    174 c    185 g    119 t
ORIGIN

  Query Match            10.0%;  Score 100.4;  DB 28;  Length 622;
  Best Local Similarity   55.8%;  Pred. No. 6.1e-13;
  Matches  191;  Conservative    0;  Mismatches  151;  Indels    0;  Gaps    0;

Qy         77 GCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTT 136

Db        610 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 551

Qy        137 GCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 196

Db        550 GCGTGCTCGGTCGCCACGTAGCGTGATCGGTCGCCACGTGGCGTGATCGGTCGCCACGTA 491

Qy        197 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 256

Db        490 GCGTGATCGGTCGCCACGTGGCGTGCTCGGTCGCCACGTAGCGTGTTCGGTCGCCACGTA 431

Qy        257 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 316

Db        430 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 371

Qy        317 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 376

Db        370 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 311

Qy        377 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 418

Db        310 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGC 269


RESULT 64
BZ216441/c
LOCUS       BZ216441                592 bp    DNA     linear   GSS 11-OCT-2002
DEFINITION  CH230-323I7.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-323I7, genomic survey sequence.
ACCESSION   BZ216441
VERSION     BZ216441.1  GI:23874799
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-323I7.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 323  row: I  column: 7
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-323I7"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      201 a    177 c    104 g    110 t
ORIGIN

  Query Match            10.0%;  Score 100.2;  DB 29;  Length 592;
```

```
  Best Local Similarity   61.1%;  Pred. No. 6.7e-13;
  Matches  162;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy        155 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 214
              |||||   |  |  |   |  ||  | || | | |||||  |  | | | |  |||
Db        394 CTGCTGCTTTTGCTGTTTGCTGGATTACTGGATTGCTGGATTAATGGATTACTGGATTAC 335

Qy        215 AGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 274
              ||  |||||| ||||| |   |   | | |||||  |||| | ||||| | | ||  |
Db        334 TGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGA 275

Qy        275 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 334
              | ||| | |||||||  | || | || || | |  |  | |||||| | | |||||  |||
Db        274 TCGCTGGATCGCTGGATTACTGGATTACTGGATTACTGGATTGCTGGATTGCTGGATTGC 215

Qy        335 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 394
              | | ||||| | | ||  | ||||| | | |||||  |||| | ||||| | | ||  |
Db        214 TGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGA 155

Qy        395 TTGCTTGCTCGCTGGCCTGCTTGCT 419
              |||||  | | |||||  |||| | |
Db        154 TTGCTGGATTGCTGGATTGCTGGAT 130


RESULT 65
BX242384/c
LOCUS       BX242384                 702 bp    DNA     linear   GSS 29-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-282O18, genomic survey sequence.
ACCESSION   BX242384
VERSION     BX242384.1  GI:28164718
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 702)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 282O18. 282O18
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .702
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-282O18"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
BASE COUNT      221 a    282 c     44 g    155 t
ORIGIN

  Query Match            10.0%;  Score 100.2;  DB 29;  Length 702;
  Best Local Similarity   50.3%;  Pred. No. 7e-13;
  Matches  272;  Conservative    0;  Mismatches  268;  Indels    1;  Gaps    1;

Qy        181 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCT 240
              | | | | |||   | | | ||   |||  |  ||   | ||| | | | |  ||
Db        701 GTTGGTTGGTTGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATG 642

Qy        241 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 300
              | | | |||   | | | | | | | |||   | | | | | | | |||   | | | |
Db        641 GATGGATGGATGGATGGATGGATGGATGGATGGATGGTTGGTTGGTTGGTTGGTTG 582

Qy        301 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 360
              | | | |||   | | | | | | | |||   | | | | | | | |||   | | | |
Db        581 GTTGGATGGATGGATGGATGGATGGTTGGTTGGATGGATGGATGGATGGATGGATGATG 522

Qy        361 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 420
              | | | |||  || ||| ||| | | | |  || ||| ||| | | | |  ||  ||
Db        521 GATGGATGGA-TGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGATGGAT 463

Qy        421 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 480
              ||  |   | | || ||  ||  ||    |  |   | | |  ||  ||   |  ||
Db        462 AGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGAT 403

Qy        481 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 540
              | | || || | || || | |   || || | || || | |   || ||   || ||
Db        402 GGATGGATGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGATGGATGGATGGAT 343

Qy        541 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 600
              | | || ||   || ||   | | || |    || ||   | | || ||   || ||
Db        342 GGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGAT 283

Qy        601 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 660
              | | ||'|    || || | |  ||| || | ||| | ||    || || | || || |
Db        282 GGATGGATGGATGGATGGTTGGTGGGTTGGTTGGCGGATAACTTGGTTGGTTGGTTGGTT 223

Qy        661 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGG 720
                |||||     ||    |  ||||| ||  | | ||| |||   ||  |  ||  |
Db        222 TGCTGGCTGGCTCACTAGCTGGCAGGCTGACTAATTGATTGATTGAGAGCAAGACAGATG 163

Qy        721 G 721
              |
Db        162 G 162


RESULT 66
BX426381
LOCUS       BX426381                1201 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX426381 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
            clone CS0DG006YM04 5-PRIME, mRNA sequence.
ACCESSION   BX426381
VERSION     BX426381.1  GI:30784544
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0AG006BG02QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DG006YM04"
                     /tissue_type="B CELLS (RAMOS CELL LINE)"
                     /cell_line="RAMOS CELL LINE"
                     /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
BASE COUNT        7 a    571 c    208 g    215 t    200 others
ORIGIN

  Query Match            10.0%;  Score 100;  DB 13;  Length 1201;
  Best Local Similarity   37.7%;  Pred. No. 8.9e-13;
  Matches  154;  Conservative   82;  Mismatches  172;  Indels    0;  Gaps    0;
```

```
Qy          12 CGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGC 71
               |||| | |     | |  |  | |   ||| | ||| ||  | | |    || |  ||
Db         244 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 303

Qy          72 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 131
               ||  ||  || |||  ||  ||  ||     | |   |   ||  ||  || |||  ||
Db         304 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 363

Qy         132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191
               ||  ||    | |   |  ||  ||  ||    | |   |   ||  ||  || |||
Db         364 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 423

Qy         192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251
               ||  ||  ||    | |   :   |   :|: :|: ||: :   :   :|:|:|: :|
Db         424 CGCGCGCGCGCGCGCGCGCKCGCGCGCKCKCKCKCGCKCKCKCKCKCKCKCKCKCKCK 483

Qy         252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
               :|: :|: :|:|:|: ||  ||: ||  || |:|: :|  :|: :|: :|:|:|: :|
Db         484 CKCKCKCKCKCKCKCKCGCGCGCKCGCGCGCGCKCKCKCKCKCKCKCKCKCKCKCKCK 543

Qy         312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
               :|: :|: :|:|:|: :|  :|: :|: :|:|:|: :|  :|: :|: :|:|:|: :|
Db         544 CKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCK 603

Qy         372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
               ||: :|: :|:|:|: :|  :|: :|: :|:|:|: :|  :|: :|:
Db         604 CGCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCTCKCKCKCK 651


RESULT 67
BX171642/c
LOCUS       BX171642                728 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-127F16, genomic survey sequence.
ACCESSION   BX171642
VERSION     BX171642.1  GI:28003347
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 728)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 127F16. 127F16
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .728
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-127F16"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
BASE COUNT      337 a    185 c     78 g    128 t
ORIGIN

  Query Match             10.0%;  Score 99.8;  DB 29;  Length 728;
  Best Local Similarity   57.7%;  Pred. No. 8.7e-13;
  Matches  198;  Conservative    0;  Mismatches  142;  Indels    3;  Gaps    1;

Qy          72 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 131
               |||  ||  || |||||  |  || |  |   |  |       || |||||| | | | ||
Db         708 TGCTTGGTTGGTCGCTTGGTTGGTCACATGGTTGGTTTGTTGCTTGCTTGTTTGGTCGCT 649

Qy         132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191
               ||  || | |   |       || || ||  |  |  |   |   |  || ||| | | |
```

```
Db         648 TGGGTGGTCACTTGGTTTGTTGCTTGGTTAATTGGTCGCTTGGTTGGTTGGTTGGTTGGT 589

Qy         192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251
               |  || ||| |  |  |   |   || |||| |||| || |   || || ||  | |
Db         588 TGGTTGGTTGGTTGGTTGGTTGGTTGCTTGCTTCCTTGGTTGGTTGGTCTGTTGTT 529

Qy         252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
               || | ||||| | | || | || ||  ||||| |||    ||||| ||||   ||
Db         528 TGGTCGCTTGTTTGATGACTTGGTTAGTTGCTTGCT---TTGCTTTGTTGCTTTGTGTGT 472

Qy         312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
               |||||||||||| | | |  || ||| ||| | ||| || |  ||| ||||| ||| |
Db         471 TGCTTGCTTGCTTGTTTGGTTGGTTGGTTGGTTGCTTGCTTAGTTGGTTGCTTGCTTGTT 412

Qy         372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 414
               || |||||| || ||| |  | |||||||| | | |   ||||
Db         411 TGGTTGCTTACTTGCTTGGTTACTTGCTTGGTTGGTAATCTGC 369


RESULT 68
BH704622/c
LOCUS       BH704622                394 bp    DNA     linear   GSS 20-FEB-2002
DEFINITION  BOMBE06TR BO_2_3_KB Brassica oleracea genomic clone BOMBE06,
            genomic survey sequence.
ACCESSION   BH704622
VERSION     BH704622.1  GI:18786296
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 394)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOMBE06TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .394
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOMBE06"
                     /clone_lib="BO_2_3_KB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       96 a    137 c    118 g     43 t
ORIGIN

  Query Match             9.9%;  Score 99.2;  DB 28;  Length 394;
  Best Local Similarity   54.0%;  Pred. No. 1e-12;
  Matches  203;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy          73 GCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 132
               || |||  | |||||    || ||||  |  |   |  | || |||| | |||||
Db         377 GCGTGCTCGGTCGCTACGAAGCGTGCTCGGTCGCTACGAAGCGTGCTCGGTCGCTACGAA 318

Qy         133 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 192
               || ||||  |  |  |    || || | | |         ||   ||  | | |||   |
Db         317 GCGTGCTCGGTCGCTACGTAGCGTGATCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 258

Qy         193 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCT 252
               || |||| | |        ||   |  | | |||   |||   |  | |||||    |
```

```
Db        257 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTT 198

Qy        253 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 312
              || |||| | |||||     ||| |||| | |||||      || |||| | |||||
Db        197 GCGTGCTCGGTCGCTACGTTGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 138

Qy        313 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 372
              || |||| | |||||     || |||| | |||||      || |||| | |||||
Db        137 GCGTGCTCGGTCGCTACGAAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 78

Qy        373 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAG 432
              || |||| | |||||     || |||| | |||||      || ||||  |  |  |
Db         77 GCGTGCTCGGTCGCTACGAAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 18

Qy        433 GGCTGCTTGCTTGCTA 448
              |  |||| | | ||||
Db         17 GCGTGCTCGGTCGCTA 2
```

```
RESULT 69
CNS05B34
LOCUS       CNS05B34                1094 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T3 end of clone
            020G19 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL329305
VERSION     AL329305.1  GI:8222927
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1094)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1094
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="020G19"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA020AD10A1~end : T3"
BASE COUNT      142 a    321 c    296 g    317 t     18 others
ORIGIN
```

```
  Query Match             9.9%;  Score 99.2;  DB 29;  Length 1094;
  Best Local Similarity   54.1%;  Pred. No. 1.3e-12;
  Matches  200;  Conservative    0;  Mismatches  170;  Indels    0;  Gaps    0;

Qy         58 ACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTG 117
              ||| |  ||  | ||||||||||||   |   |||||| |||       | | ||  | |||
Db        141 ACAGCCTGTCTGTCTGCCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGTCTGTCTGTCTG 200

Qy        118 CTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGG 177
                ||   |   |||||  ||       | | ||  | ||||  ||       | | ||  |
Db        201 TCTGTCTGTCTGCCTGTCTGTCTGTCTGTCTGTCTGTCTGCCTGTCTGTCTGTCTGTCTG 260

Qy        178 CCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGG 237
              ||||  ||   |   |||||  |||       | | ||  | |||   || |||  ||  |
Db        261 TCTGCCTGTCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGTCTG 320

Qy        238 GCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 297
              ||||  |   |||||  ||  ||   ||  |||||  ||  |||  ||  |||||  ||
Db        321 TCTGCCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTG 380

Qy        298 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 357
                ||   ||  |||||| ||  |||  ||  | |||  |||  |||  ||  | | |  ||
Db        381 TCTGTCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTGTCTGCCTGCCTGCCTGTCNGTCTG 440

Qy        358 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 417
                |||  ||  |||||  |||  ||   |  | ||||  |||  ||   ||  | |||
Db        441 TCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGTCTGCCTGTCTGTAGN 500

Qy        418 CTGAGGGGAC 427
              | ||   |||
Db        501 CAGATATGAC 510
```

```
RESULT 70
BH981983/c
LOCUS       BH981983                787 bp    DNA     linear   GSS 07-OCT-2002
DEFINITION  ode73b07.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
            sequence.
ACCESSION   BH981983
VERSION     BH981983.1  GI:23513062
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 787)
  AUTHORS   Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash
            ,W., Rabinowicz,P.D. and Wilson,R.K.
  TITLE     Whole genome shotgun reads from Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: ode73  row: b  column: 07
            Seq primer: -21UPpOT forward
            Class: shotgun
            High quality sequence start: 70
            High quality sequence stop: 271.
FEATURES             Location/Qualifiers
     source          1. .787
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3712"
                     /clone_lib="B.oleracea002"
                     /note="Vector: pOTw13; Whole genome shotgun library from
                     flowering buds. DNA was purified  from a crude nuclear
                     prep using Brassica oleracea TO1000DH3 buds  provided by
                     Thosmas Osborn at the University of Wisconsin. Genomic
                     DNA was provided by Pablo Rabinowicz (CSHL) and the
                     shotgun library  prepared at Washington University Genome
                     Sequencing Center."
```

```
BASE COUNT     230 a    276 c    184 g     95 t      2 others
ORIGIN

  Query Match             9.9%;  Score 99;  DB 28;  Length 787;
  Best Local Similarity   55.8%;  Pred. No. 1.4e-12;
  Matches  189;  Conservative    0;  Mismatches  150;  Indels    0;  Gaps    0;

Qy         74 CCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTG 133
              | |||  | |||  |    || || |  |  |   |  || |||| | ||||||    |
Db        480 CGTGCTCGGTCGTCGCGTAGCGTGGTCGGTCGCTGGGTAGCGTGCTCGGTCGCTGCGTAG 421

Qy        134 CTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGG 193
              | ||||  |  |       || |||| | |    |  ||   ||  | | |||   | |
Db        420 CGTGCTCGGTCGCAGCGTAGCGTGCTCGGTCGCAGCGTAGCGTGCTCGGTCGCTGCGTAG 361

Qy        194 CCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTG 253
              | |||| | |    |   ||   |   | | |||   |||   |   | ||||||    |
Db        360 CGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAG 301

Qy        254 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 313
              | |||| | ||||||    || |||| | ||||||    || |||| | ||||||    |
Db        300 CGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAG 241

Qy        314 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 373
              | |||| | ||||||    || |||| | ||||||    || |||| | ||||||    |
Db        240 CGTGCTCGGTCGCTCCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAG 181

Qy        374 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 412
              | |||| | ||||||    || |||| | |||||  | |
Db        180 CGTGCTCGGTCGCTGCGTAGCGTGCTCGATCGCTCCCGT 142


RESULT 71
CA056594/c
LOCUS       CA056594                734 bp    mRNA    linear   EST 04-MAR-2003
DEFINITION  ssalrgb536229 mixed_tissue Salmo salar cDNA, mRNA sequence.
ACCESSION   CA056594
VERSION     CA056594.1  GI:24386837
KEYWORDS    EST.
SOURCE      Salmo salar (Atlantic salmon)
  ORGANISM  Salmo salar
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
REFERENCE   1  (bases 1 to 734)
  AUTHORS   GRASP Consortium, Davidson,W.S., Koop,B.F. and
            http://web.uvic.ca/cbr/grasp.
  TITLE     A survey of Salmo salar transcripts from high complexity cDNA
            libraries
  JOURNAL   Unpublished
COMMENT     Contact: Koop BF
            Centre for Biomedical Research
            University of Victoria
            PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
            Tel: 250 472 4067
            Fax: 250 472 4075
            Email: bkoop@uvic.ca
            Genome Sciences Centre, BC Cancer Agency
            cDNA preparation, sequencing and bioinformatics:
            Y Butterfield, R Kirkpatrick, J Asano, N Girn, R Guin, D Lee,
            S Lee, T Olson, P Pandoh, A Prahbu, D Smailus, L Spence, J Stott,
            S Taylor, G Yang, J Schein, S Jones and M Marra.
FEATURES             Location/Qualifiers
     source          1. .734
                     /organism="Salmo salar"
                     /mol_type="mRNA"
                     /strain="McConnell"
                     /db_xref="taxon:8030"
                     /clone_lib="mixed_tissue"
                     /note="Vector: pCMVsport6; Library Creator:  Research
                     Genetics ; Atlantic salmon tissue contributors: Carlo
                     Biagi, Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.),
                     Simon Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery
                     (Crofton, B.C.), Rachel Roper (University of Victoria)"
BASE COUNT     276 a    187 c    162 g    109 t
ORIGIN

  Query Match             9.9%;  Score 98.6;  DB 14;  Length 734;
  Best Local Similarity   48.9%;  Pred. No. 1.6e-12;
  Matches  295;  Conservative    0;  Mismatches  304;  Indels    4;  Gaps    1;

Qy        131 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGC 190
              |||  |||   || ||||||  | ||||  ||       | .|    | |||  ||   ||
Db        733 CTGACTGCAATGGTGACAGTCTGTCTGCCTGTCTGACTGCTATGGTGTCTGTCTGTCTGC 674

Qy        191 TGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGC 250
              |    ||   |        | | ||  | |||   |  |||||    | | |   |   |
Db        673 TATGGTGACAGTCTGTCTGCCTGTCTGTCTGTCTGACTGCTATGGTGACAGTCTGTCTGT 614

Qy        251 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 310
              |||  ||  ||   |||    ||   |  ||   |   | |||  ||  ||   |   |
Db        613 CTGTCTGTCTGACTGCTATGGTGACAGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 554

Qy        311 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 370
              |||  ||  ||   ||     ||   |  ||   ||  | |||  ||||      |   |
Db        553 CTGTCTGTCTGACTGCAATGGTGACAGTCTGTCTGCCTGTCTGACTGCTATGGTGTCTGT 494

Qy        371 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 430
              |||  ||||     |   | |||  ||| ||   |   | |||  ||||  || ||||||
Db        493 CTGTCTGCTATGGTGACAGTCTGTCTGCCTGTCTGTCTGTCTGACTGCTATGGTGACAGT 434

Qy        431 AGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGC 490
              |||  ||  ||||| || | | ||  | ||  ||| |    ||| |||  | ||||
Db        433 ----CTGTCTGTCTGCTATGGTGTCTGTCTGTCTAACTGCTATGGTGACAGTCTGTCTGC 378

Qy        491 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 550
              | |||   ||  | |||||  | ||  |||  | |||   | |||   |   |
Db        377 CTGTCTGTCTGTCTGACTGCTATGGTGACAGTCTGTCTGTCTGTCTGTCTGACTGCTATG 318

Qy        551 TGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCC 610
              | | |   ||  | ||    | |||| |    | | |   | ||||  ||  | || |
Db        317 GTGACAGTCTGTCTGTCTGTATGACTGCAATGGTGACAGTCTGTCTGCCTGTCTGACTGC 258

Qy        611 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 670
              |  | ||   ||  | |||||  | ||  |||  | ||||  | |||   ||  | ||||
Db        257 TATGGTGTCTGTCTGTCTGCTATGGTGACAGTCTGTCTGCCTGTCTGTCTGTCTGACTGC 198

Qy        671 TGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGC 730
              | ||  |||  | |||||||   |  ||||   ||  ||   ||  |    ||
Db        197 AATGGTGACAGTCTGTCTGCCTGTCTGTCTGCTATGGTGTCTGTCTGTCTGCTATGGTGT 138

Qy        731 CTG 733
              |||
Db        137 CTG 135


RESULT 72
BH115707/c
LOCUS       BH115707                310 bp    DNA     linear   GSS 19-JUL-2001
DEFINITION  RPCI-24-298L24.TJ RPCI-24 Mus musculus genomic clone RPCI-24-298L24
            , genomic survey sequence.
ACCESSION   BH115707
VERSION     BH115707.1  GI:14955201
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 310)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished
COMMENT     Other_GSSs: RPCI-24-298L24.TV
            Contact: Shaying Zhao
```

```
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 298  row: L  column: 24
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .310
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-298L24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /clone_lib="RPCI-24"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT       97 a     75 c    134 g      4 t
ORIGIN

  Query Match             9.8%;  Score 98.4;  DB 28;  Length 310;
  Best Local Similarity   61.9%;  Pred. No. 1.5e-12;
  Matches  156;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        218 GCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 277
              | |||  || |||  |   | ||||  ||  |||||| ||| |||  ||  |||||||||
Db        308 GTTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCTTG 249

Qy        278 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 337
                |||  |   ||||||  ||| | ||  |||  |   ||  |   ||     |||||  ||
Db        248 TCTGCCTGTCTGCCTGCCTGCCTCCTGTCTGCCTGTCTGCCCGCCTGATCTGCCTGTCTG 189

Qy        338 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 397
              | |||  ||  |||||  ||| |||  ||  ||||||| ||| |||  ||  ||||||| ||
Db        188 CCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTG 129

Qy        398 CTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGG 457
              | |||  ||  |||||| |||      | | ||  | |||| ||| |||      | | |
Db        128 CCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTG 69

Qy        458 TACGCCTGCCTG 469
              |  |||||||||
Db         68 TCTGCCTGCCTG 57


RESULT 73
BI688458/c
LOCUS       BI688458                745 bp    mRNA    linear   EST 18-SEP-2001
DEFINITION  603316395F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5356246 5',
            mRNA sequence.
ACCESSION   BI688458
VERSION     BI688458.1  GI:15651087
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 745)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11905  row: n  column: 23
            High quality sequence stop: 650.
FEATURES             Location/Qualifiers
     source          1. .745
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5356246"
                     /sex="female, virgin"
                     /tissue_type="infiltrating ductal carcinoma"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam6"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      280 a    140 c    187 g    138 t
ORIGIN

  Query Match             9.8%;  Score 97.6;  DB 12;  Length 745;
  Best Local Similarity   68.5%;  Pred. No. 2.8e-12;
  Matches  178;  Conservative    0;  Mismatches   79;  Indels    3;  Gaps    3;

Qy        218 GCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 277
              | | | || ||||| |   |  |||| ||| || ||||||| || | | | || ||||||
Db        325 GGTTCAGGTTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTG-TTGTTTGCTTGCTTG 267

Qy        278 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 337
              ||||||  ||| || |||||||||||||  ||| || ||||||||  || | | |  ||||||
Db        266 CTTGCTTGCT-GCTTGCTTGCTTGCTTGCTTGCTTGCTGCCT-GTTCGTTTGCTTG 209

Qy        338 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 397
              ||||||  ||| || |||||||||||||  ||| || ||||||||||||  ||| ||||| | |
Db        208 CTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGTTCG 149

Qy        398 CTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGG 457
              |||||  ||| || ||||  |||   |     | |  ||| || |    |   || |  |
Db        148 TTTGCTTGCTTGCTTGCTGCCTGTTCGTTTGCTTGCTCTGTTTACCAAGTTCTGGTATAG 89

Qy        458 TACGCCTGCCTGATGGCTTG 477
              | |     | |||    |||
Db         88 TGCAATGCCATGAACCTTTG 69


RESULT 74
BX245766/c
LOCUS       BX245766                550 bp    DNA     linear   GSS 13-MAR-2003
DEFINITION  Danio rerio genomic clone DKEY-231B6, genomic survey sequence.
ACCESSION   BX245766
VERSION     BX245766.1  GI:28168100
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
```

```
COMMENT     This sequence was generated from the T7 end of BAC 231B6. 231B6 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .550
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-231B6"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
BASE COUNT      248 a    230 c     23 g     49 t
ORIGIN

  Query Match             9.7%;  Score 97.4;  DB 29;  Length 550;
  Best Local Similarity   54.6%;  Pred. No. 2.9e-12;
  Matches  194;  Conservative    0;  Mismatches  161;  Indels    0;  Gaps    0;

Qy        208 GGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCT 267
              ||   || ||| || | | ||| | |   |   | | | | |  || ||| | | ||| |
Db        490 GGTTGGTTGGGTTGGTTGGTTGGTTGGTCGATCGGTTGGTTGGTTGGTTGGTCGATCGGT 431

Qy        268 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 327
                |  || ||| ||| ||| | |  || ||| ||| | | |||   | ||| ||| | | |
Db        430 TGGTTGGTTGGTTGGTCGATCGGTTGGTTGGTTGGTTGTTGGTTAGTTTGATTGGTTGGT 371

Qy        328 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 387
                |  || ||| | | ||  | |  || ||| ||| | | | |  || ||| ||| | | |
Db        370 TGGTTGGTTGGTGGATCAGTTGGTTGGTTGGTTGTTGGTTAGTTTGATTGGTTGGTTGGT 311

Qy        388 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCT 447
                |  || |||  || ||| | |  || ||| |  |  ||  |   || || ||| ||| |
Db        310 TGGTTGGTTGGATGATCGGTTGGTTGGTTGGTTGGATGATCGGTTGGTTGGTTGGTTGGT 251

Qy        448 AAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGC 507
                |  |   |   |  ||  |||| | ||| | |   || ||   |   | | |   ||
Db        250 TGGTTGGTTGGTTGGTTGGATGATCGGTTGGTTGGTTGGTTGGATGATCGGTTGGTTGGT 191

Qy        508 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGT 562
              || | || || | |   || || | || ||   |   || || | |  ||  | |
Db        190 TGGTTGGATGATCGGTTGGTTGGTTGGTTGGATGGTTGGATGGTTGATTGGGATT 136


RESULT 75
BG561076
LOCUS       BG561076                393 bp    mRNA    linear   EST 10-APR-2001
DEFINITION  EtESTed91c07.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
            tenella cDNA 5', mRNA sequence.
ACCESSION   BG561076
VERSION     BG561076.1  GI:13590074
KEYWORDS    EST.
SOURCE      Eimeria tenella
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 393)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 282.
FEATURES             Location/Qualifiers
     source          1. .393
                     /organism="Eimeria tenella"
                     /mol_type="mRNA"
                     /strain="LS18"
                     /db_xref="taxon:5802"
                     /dev_stage="Sporozoite stage"
                     /lab_host="SOLR"
                     /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                     ; Sporozoites were obtained from in vitro sporulated and
                     excysted oocysts of E.tenella grown in chickens. cDNA was
                     synthesized from poly mRNA using an oligo-dT primer
                     containing a XhoI site. Following second strand synthesis,
                     EcoRI adapters were ligated to the cDNA and products were
                     size-selected on sephacryl S500. The cDNA were ligated to
                     EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were
                     converted to phagemids by mass excision using ExAssist
                     helper phage and E.coli SOLR cell (Stratagene). Clones
                     were selected by negative hybridization against a pool of
                     overrepresented ESTs (N>=10,from 1682 previous reads).
                     Insert sizes range from 1.2-2.9Kb. The library may contain
                     a small percentage of host or bacterial contaminents."
BASE COUNT       81 a    119 c    120 g     73 t
ORIGIN

  Query Match             9.7%;  Score 97.2;  DB 10;  Length 393;
  Best Local Similarity   62.2%;  Pred. No. 2.9e-12;
  Matches  153;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy        221 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 280
              ||||||| |||||   |  |||   || ||  ||| | | ||| ||| ||  ||| |||
Db        123 GCTGGCTAGCTAGCTAGCTTGCCATCTAGCTAGCTGGTTGGCTAGCTAGCTGGCTAGCTA 182

Qy        281 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 340
              ||  ||   || |||  || ||| ||| ||| ||  ||| ||| | ||||   || | |
Db        183 GCCAGCGAACCGGCTGCCTGGCTGGCTAGCCAGCGAGCTAGCTGGGTGGCTAACTGGATG 242

Qy        341 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
              ||| ||| ||| ||  ||| ||| | ||||  ||| | | ||| ||||||  ||    |
Db        243 GCTGGCTAGCCAGCGAGCTAGCTGGGTGGCTAGCTGGATGGCTAGCTGGCTAGCCACATA 302

Qy        401 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTAC 460
              ||| ||  ||  ||| ||| ||  | ||   | || ||| ||  ||||  |||   |
Db        303 GCTAGCCAGCTGGCTGGCTCAGCAGCCACCTGAGCAGCTAGCCAGCTACTGGGGGAGGCA 362

Qy        461 GCCTGC 466
              ||  ||
Db        363 GCGCGC 368


RESULT 76
CD347167/c
LOCUS       CD347167                579 bp    mRNA    linear   EST 29-MAY-2003
DEFINITION  EtESTef11c07.y1 Eimeria tenella M5-6 Excised cDNA Eimeria tenella
            cDNA 5' similar to TR:Q20000 Q20000 SIMILARITY WITH VON WILLEBRAND
            FRACTOR. [1] ;, mRNA sequence.
ACCESSION   CD347167
VERSION     CD347167.1  GI:31138788
KEYWORDS    EST.
SOURCE      Eimeria tenella
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 579)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
```

```
  JOURNAL   Unpublished
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 564.
FEATURES             Location/Qualifiers
     source          1. .579
                     /organism="Eimeria tenella"
                     /mol_type="mRNA"
                     /db_xref="taxon:5802"
                     /dev_stage="LS18"
                     /lab_host="E.coli DH10B (GeneHog, Invitrogen,Inc)"
                     /clone_lib="Eimeria tenella M5-6 Excised cDNA"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Sporozoites were obtained from in vitro sporulated
                     and excysted oocysts of E.tenella grown in chickens. cDNA
                     was synthesized from poly mRNA using an oligo-dT primer
                     containing a XhoI site. Following second strand synthesis,
                     EcoRI adapters were ligated to the cDNA and products were
                     size-selected on sephacryl S500. The cDNA were ligated to
                     EcoRI/XhoI prepared lambda ZapII (Stratagene). The primary
                     library was mass excised using ExAssist helper phage
                     (Stratagene). The phagemids were preciptated with PEG 8000
                     , extracted with phenol/chloroform and electroporated into
                     DH10B cells. The library may contain a small percentage of
                     host or bacterial contaminants. Library materials provided
                     by: Paul Liberator, Merck Research Labs Library
                     constructed by: Paul Liberator, Merck Research Labs."
BASE COUNT      132 a    171 c    141 g    135 t
ORIGIN

  Query Match             9.7%;  Score 97.2;  DB 14;  Length 579;
  Best Local Similarity   59.4%;  Pred. No. 3.3e-12;
  Matches  183;  Conservative    0;  Mismatches  123;  Indels    2;  Gaps    1;

Qy        220 TGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 279
              ||   || |||  | |||  |||  ||  || ||| |||||||  ||  || ||| |||
Db        573 TGGACGCATGCATGCAGGCTTGCATGCATGCATGCGTGCTTGCATGCATGCATGCATGCA 514

Qy        280 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC--GCTGGCCTGCTTG 337
              | |  ||| || ||| ||| |||  ||| || ||| |||  || |  ||  ||.||||||
Db        513 TCCAGGCTTGCATGCATGCATGCGTGCTTGCATGCATGCAGGCACATGCATGCATGCTTG 454

Qy        338 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 397
              | |||  ||  || ||| |||||||  ||  || |||||| |||  ||  || ||| ||
Db        453 CGTGCATGCATGCATGCGTGCTTGCATGCATGCATGCTTGCGTGCATGCATGCTTGCATG 394

Qy        398 CTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGG 457
              | |||   | |  |||||||      |   | | || ||||||| |||      |
Db        393 CATGCATACATGGGTGCTTGCATGCATGCATGCACGGGTGCTTGCATGCATGCACGCGTT 334

Qy        458 TACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTG 517
               |||| ||| ||   |  ||   | |  | | || |  ||| |   |  |||  |  ||
Db        333 CACGCATGCATGGATGGGTGCGTGGACAGGTCCTTGCATGCATGCGCGCATGCATGCATG 274

Qy        518 CTAGTAGG 525
              || | | |
Db        273 CTTGCATG 266


RESULT 77
AZ437267/c
LOCUS       AZ437267                477 bp    DNA     linear   GSS 03-OCT-2000
DEFINITION  1M0225L23F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0225L23 F, genomic survey sequence.
ACCESSION   AZ437267
VERSION     AZ437267.1  GI:10561280
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
\           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 477)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0225  row: L  column: 23
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 477.
FEATURES             Location/Qualifiers
     source          1. .477
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0225L23"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      181 a     91 c    113 g     92 t
ORIGIN

  Query Match             9.7%;  Score 96.8;  DB 28;  Length 477;
  Best Local Similarity   74.4%;  Pred. No. 3.8e-12;
  Matches  122;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy        229 GCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 288
              | || || |  ||   |  | |||  || ||| | | |  || |||||||||  |||
Db        164 GATAATACGTTTGTCTGTCTGTCTGTCTGTTTGTTTGTTTGTTTGCTTGCTTGCTT 105

Qy        289 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 348
              ||||||||||||||| ||  || ||||||||||  ||| || |||||| |||| |||
Db        104 GCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTT 45

Qy        349 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 392
              || ||| |||||||| ||| ||||||||||||||| ||  ||||
```

```
Db        44 GCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCCT 1


RESULT 78
CB811482
LOCUS       CB811482                404 bp    mRNA    linear   EST 16-MAY-2003
DEFINITION  AMGNNUC:MRPE4-00166-A1-A mrpe4 (10380) Rattus norvegicus cDNA clone
            mrpe4-00166-a1 5', mRNA sequence.
ACCESSION   CB811482
VERSION     CB811482.1  GI:29934380
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 404)
  AUTHORS   Amgen EST Program.
  TITLE     Amgen Rat EST Program
  JOURNAL   Unpublished
COMMENT     Contact: Dan Fitzpatrick
            Amgen, Inc
            One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
            Tel: 805 447-4881
            Plate: 00166  row: a  column: 1.
FEATURES             Location/Qualifiers
     source          1. .404
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="mrpe4-00166-a1"
                     /tissue_type="placenta embryo"
                     /clone_lib="mrpe4 (10380)"
                     /note="Vector: pSPORT1; Site_1: Sal1; Site_2: NotI;
                     placenta embryo day 17"
BASE COUNT       36 a    154 c    102 g    110 t      2 others
ORIGIN

  Query Match             9.7%;  Score 96.6;  DB 14;  Length 404;
  Best Local Similarity   60.5%;  Pred. No. 4.1e-12;
  Matches  159;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy        247 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 306
              ||  |||| ||| |||   ||  |||||| ||| |||  ||  |||||| ||| |||  ||
Db        142 TGCTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGC 201

Qy        307 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 366
                |||||| ||| |||  ||  |||||| ||  |||  |   |||||  ||| |||. ||
Db        202 CTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGCCTGC 261

Qy        367 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 426
                |||||  ||  |||  |   ||||  | |||  ||   ||  ||||||  |||       |
Db        262 CTCCCTGCCTGTCTGCCTGTCTGCCTCCCTGCCTGTCTGCCTGCCTGCCTGCCTCCCTGC 321

Qy        427 CAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGG 486
              |  ||   |  ||||  |||  |||    |   |  |  |     ||||||||    ||  ||    |    |
Db        322 CTGTCTGCCTGCCTGCCTGCCTGGCTGCCTGCCTCCCTGCCTGTCTGCCTGCCTGCCTGC 381

Qy        487 CTGCTGGGCTGCTAGTAGGGCTG 509
              |||    |  ||||   |    | |||  .
Db        382 CTGGCTGCCTGCCTGGCTGCCTG 404


RESULT 79
DR42N9S/c
LOCUS       DR42N9S                 490 bp    DNA     linear   GSS 22-NOV-2002
DEFINITION  Danio rerio genomic clone DKEY-42N9, genomic survey sequence.
ACCESSION   AL981355
VERSION     AL981355.1  GI:25183648
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   Humphray,S.J., Huckle,E. and Hunt,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2002) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 42N9. 42N9 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene.
            Further details: http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .490
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-42N9"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
BASE COUNT      176 a    170 c     11 g    133 t
ORIGIN

  Query Match             9.6%;  Score 96.2;  DB 29;  Length 490;
  Best Local Similarity   53.3%;  Pred. No. 5.3e-12;
  Matches  203;  Conservative    0;  Mismatches  178;  Indels    0;  Gaps    0;

Qy        180 TGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGC 239
              || |||. | | | |  || ||| |     ||  |   |  || | | ||| | |   |
Db        482 TGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGAT 423

Qy        240 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 299
              || | | |  || ||| ||| | | | |  || ||| ||| | | | |  || ||| |
Db        422 TGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGAT 363

Qy        300 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 359
              || | | |  || ||| ||| | | | |  || ||| ||| | | | |  || ||| |
Db        362 TGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGAT 303

Qy        360 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              || | | |  || ||| ||| | | | |  || ||| ||| | |   | | | | | |
Db        302 TGATTGATTGATTGATTGATTGATTGATTGATTGATTGATGGGCAGACGGATGGAT 243

Qy        420 GAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGAT 479
              |   || |||  || | | | | | | |    ||| ||   |   |  ||||| | |||
Db        242 GGACGGGCAGACGGACGGATGGATGGATGGATGGATGGATGGATGGATGGATGGAT 183

Qy        480 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 539
              |  ||   | |||   | | |  ||   | |||   | | |  ||   | |||   | |
Db        182 GGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGAT 123

Qy        540 AGTAGGGCTGCTGGGCTGCTA 560
              |  ||   | |||   | ||
Db        122 GGATGGATGGATGGATGGATA 102


RESULT 80
AI798117
LOCUS       AI798117                615 bp    mRNA    linear   EST 20-DEC-1999
DEFINITION  wh81d08.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2387151 3'
            similar to contains Alu repetitive element;contains element TAR1
            repetitive element ;, mRNA sequence.
ACCESSION   AI798117
VERSION     AI798117.1  GI:5363589
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 615)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
```

```
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
            M.D., Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1082   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 458.
FEATURES             Location/Qualifiers
     source          1. .615
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2387151"
                     /tissue_type="B-cell, chronic lymphotic leukemia"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_CLL1"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTGCTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT       77 a    229 c    157 g    152 t
ORIGIN

  Query Match             9.6%;  Score 96.2;  DB 9;  Length 615;
  Best Local Similarity   54.7%;  Pred. No. 5.6e-12;
  Matches  215;  Conservative    0;  Mismatches  173;  Indels    5;  Gaps    1;

Qy        246 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 305
              | |  || || || |     |||| || || || | |   | || || || || | |
Db         69 CCGAGCTCCTGGCCTCGCATCTGGGCTCCTGGCCTCCCACCCGGGCTCCTGGCCTCCCAC 128

Qy        306 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
              | || || ||    | |  || || || || |  | |   |||| || || || | |
Db        129 CCGGGCTCCTGAGCTCCCAGCCGGGCTCCTGGGCTTCCATCTGGGCTCCTGGCCTCCCAT 188

Qy        366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAG--- 422
              || || || || |     |||| || || || | |     |||| || || ||
Db        189 GAGGGCTCCTGGCCTCACATCTGGGCTCCTGGCCTCACATCTGGGCTCCTGGCCTCACAT 248

Qy        423 --GGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 480
                |||   | |  || |   || |      || |     ||| | ||  |||| | | |
Db        249 CTGGGCTCCTTGCCCTCCCACCTGGGCTCCTGGGCTCCCAGCCAGGCTCCTGGCCTCACA 308

Qy        481 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 540
              |||| |||| || | |    |||| |||| ||   |   |||| |||| ||   |
Db        309 TCTGGGCTCCTGGCCTCCCATCCAGGCTCCTGGCCTTGAATCTGGGCCCCTGGCCTCACA 368

Qy        541 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 600
              |||| |||| || | |  |||||||||| || | |   ||| | |||| ||| |
Db        369 TCTGGGCTCCTGGCCTCCCATCAGGGCTCCTGGACTCCCATCTGGGATCCTGGCCTGGAA 428

Qy        601 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGG 633
              |||| ||| || | |    |||| || |
Db        429 TCCAGGCTCTTGGCCTCCCATCCAGGCTCCTAG 461


RESULT 81
BZ224494
LOCUS       BZ224494               1039 bp    DNA     linear   GSS 11-OCT-2002
DEFINITION  CH230-446O21.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-446O21, genomic survey sequence.
ACCESSION   BZ224494
VERSION     BZ224494.1  GI:23882852
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1039)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-446O21.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 446  row: O  column: 21
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .1039
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-446O21"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      265 a    229 c    278 g    267 t
ORIGIN

  Query Match             9.6%;  Score 95.8;  DB 29;  Length 1039;
  Best Local Similarity   52.3%;  Pred. No. 8e-12;
  Matches  308;  Conservative    0;  Mismatches  272;  Indels    9;  Gaps    4;

Qy         89 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 148
              | | |||  ||||| ||||||  ||   ||      ||  | | |||  ||||| |||
Db        456 GTCAGCTGTCTGAGATGACAGTCAGCTGTCTGAGATGCCAGTCGGCTGTCTGAGATGAC- 514

Qy        149 GTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 208
              |   | | |||  |||||  | ||||  ||   ||      ||| | | |||  |||||
Db        515 GACAGTCAGCTCTCTGAGATGCCAGTCAGCTGTCTGAGATGCTAGTCAGCTGTCTGAGAT 574

Qy        209 GACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 268
              | |||| ||    |||  ||| |||  |    ||      | || |   |  |||  |||
Db        575 GCCAGTCGGCTGTCTGAGATGCCAGTCAGCTGTCTGAGATGACTACAGTCA-GCTGTCTG 633

Qy        269 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 328
              ||   |  |||  |||  ||| || | |||  |||  ||   |   |||  |||  |||
Db        634 AGATGACAGTCGGCTGTCTGAGATGCCTGTTGGCTGTCTGAGATGACAGTCAGCTGTCTG 693

Qy        329 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 388
              |||  |  |||| |||  |||  |  |||  |||  |||  |  |||  |||
Db        694 AGATGCCAGTCAGCTCTCTGAGATGCCAGTCAGCTGTCTGAGATGCCAGTCAGCTGTCTG 753

Qy        389 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTA 448
                ||   |  |||  |||  |||  | ||  |  .  | | ||| | |||  ||
```

```
Db      754 AGATGACAGTCGGCTGTCTGAGATGCCAGTTGGCTGTCTGAGATGCCTGTTGGCTGTCTG 813

Qy      449 AGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 508
            ||  ||| ||    | ||    ||   || |||   |    ||| | ||  ||| ||
Db      814 AGATGACAGT----CAGCTGTCTGAGATGCTAGCCAGCTGTCTGAGATGACAGTTGGCTG 869

Qy      509 GCTGGGCTGCTAGTAGGGCTGCTGGGCTG---CTAGTAGGGCTGCTGGGCTGCTAGTAGG 565
             ||| | ||| ||| |    ||| | ||     ||| |    ||| | ||| ||| ||
Db      870 TCTGAGATGCCAGTCAGCTGTCTGAGATGACTACAGTCAGCTGTCTGAGATGCCAGTCGG 929

Qy      566 GCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGG 625
                ||| | ||| ||| |    ||| | ||| ||| |    ||| | ||| ||| |
Db      930 CTGTCTGAGATGCCAGTCAGCTGTCTGAGATGCCAGTCAGCTGTCTGAGATGCCAGTCAG 989

Qy      626 GCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 674
                ||| | ||  ||| ||    ||| | ||| ||| ||    ||| |
Db      990 CTGTCTGAGATGACAGTCGGTTGTCTGAGATGCCAGTTGGCTGTCTGAG 1038


RESULT 82
AZ857141/c
LOCUS       AZ857141               421 bp    DNA     linear   GSS 21-FEB-2001
DEFINITION  2M0161O23R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0161O23 R, genomic survey sequence.
ACCESSION   AZ857141
VERSION     AZ857141.1  GI:13049041
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 421)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0161  row: O  column: 23
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 421.
FEATURES             Location/Qualifiers
     source          1. .421
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0161O23"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      150 a     88 c     115 g     68 t
ORIGIN

  Query Match           9.6%;  Score 95.6;  DB 28;  Length 421;
  Best Local Similarity   73.5%;  Pred. No. 7e-12;
  Matches  122;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy      227 TTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 286
            ||| |  |   | |  |    ||  || |||||||||| ||  || ||| |||||||  ||
Db      194 TTGGTTCTGTTGTTATTGTTGGGTGTGTTTGCTTGCTTGCCTGCTTGCCTGCTTGCCTGC 135

Qy      287 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 346
            | |||||||||| |||| ||  || ||| |||||||  ||| ||||||||| |||| ||
Db      134 TTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGC 75

Qy      347 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 392
              || ||| |||||||| ||| ||||||||||||||| ||  ||||
Db       74 CTGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCCT 29


RESULT 83
BZ425593/c
LOCUS       BZ425593               839 bp    DNA     linear   GSS 13-DEC-2002
DEFINITION  BONIO62TF BO_1.6_2_KB_tot Brassica oleracea genomic clone BONIO62,
            genomic survey sequence.
ACCESSION   BZ425593
VERSION     BZ425593.1  GI:26667048
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 839)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BONIO62TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .839
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BONIO62"
                     /clone_lib="BO_1.6_2_KB_tot"
                     /note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared
                     total DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      208 a     227 c     202 g     202 t
ORIGIN

  Query Match           9.5%;  Score 95;  DB 29;  Length 839;
  Best Local Similarity   52.7%;  Pred. No. 1.2e-11;
  Matches  206;  Conservative    0;  Mismatches  185;  Indels    0;  Gaps    0;

Qy       73 GCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 132
```

```
Db        838 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTATGTA 779

Qy        133 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 192
Db        778 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTATGTAGCGTGCTCGGTCGCTACGTA 719

Qy        193 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCT 252
Db        718 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 659

Qy        253 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 312
Db        658 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 599

Qy        313 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 372
Db        598 GCGTGCTCGGTCGATACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 539

Qy        373 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAG 432
Db        538 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTATGTAGCGTGCTCGGTCGCTACGTA 479

Qy        433 GGCTGCTTGCTTGCTAAGGGGACGGTACGCC 463
Db        478 GCGACCGTGCTTTCTTAAAAATCGATACGAC 448


RESULT 84
AK089803/c
LOCUS       AK089803               2933 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus activated spleen cDNA, RIKEN full-length enriched
            library, clone:F830023K21 product:unclassifiable, full insert
            sequence.
ACCESSION   AK089803
VERSION     AK089803.1  GI:26105561
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2933)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
            Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
            Trust/MRC building Addenbrookes Hospital Cambridge) whose
            assistance we gratefully acknowledge.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .2933
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="NOD"
                     /db_xref="FANTOM_DB:F830023K21"
                     /db_xref="taxon:10090"
                     /clone="F830023K21"
                     /tissue_type="activated spleen"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
     misc_feature    1. .2933
                     /note="unclassifiable"
BASE COUNT      867 a    625 c    616 g    825 t
ORIGIN
```

```
  Query Match          9.5%;  Score 95;  DB 11;  Length 2933;
  Best Local Similarity  63.7%;  Pred. No. 1.6e-11;
  Matches  160;  Conservative    0;  Mismatches   90;  Indels    1;  Gaps    1;

Qy        169 GACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTT 228
              | | ||  |  ||| |||  ||  |||||| |||      | | |   | |||||||||
Db       1525 GGCTGTCTGTTTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCTGGCTG 1466

Qy        229 GCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 288
              || ||  | ||||  ||  |||||||||  ||   | | |||||| ||| |||  |
Db       1465 TCT-GTTTGCCTGCCTGCCTGCCTGCTTGGCTGTCTGTTTGCCTGCCTGCCTGCCTGGCT 1407

Qy        289 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 348
              | ||| |||| |||  ||  ||||||  || ||||  ||  ||||||  ||  ||    | |
Db       1406 GTCTGTTTGCCTGCCTGCCTGCCTGCCTGTTTGCCTGCCTGCCTGCCTGGCTGTCTGTTT 1347

Qy        349 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 408
              |||||  |||||||  ||  | |||  || ||||  ||  ||||||  ||| | |  ||
Db       1346 ACCTGCCTGCTTGCCTGCCTGGCTGTCTGTTTGCCTGCCTGCCTGCCTGCCTTCCTGCCT 1287

Qy        409 GCCTGCTTGCT 419
              |||||| ||||
Db       1286 GCCTGCCTGCT 1276


RESULT 85
CA054386
LOCUS       CA054386                580 bp    mRNA    linear   EST 04-MAR-2003
DEFINITION  ssalrgb517217 mixed_tissue Salmo salar cDNA, mRNA sequence.
ACCESSION   CA054386
VERSION     CA054386.1  GI:24384629
KEYWORDS    EST.
SOURCE      Salmo salar (Atlantic salmon)
  ORGANISM  Salmo salar
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
REFERENCE   1  (bases 1 to 580)
  AUTHORS   GRASP Consortium, Davidson,W.S., Koop,B.F. and
            http://web.uvic.ca/cbr/grasp.
  TITLE     A survey of Salmo salar transcripts from high complexity cDNA
            libraries
  JOURNAL   Unpublished
COMMENT     Contact: Koop BF
            Centre for Biomedical Research
            University of Victoria
            PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
            Tel: 250 472 4067
            Fax: 250 472 4075
            Email: bkoop@uvic.ca
            Genome Sciences Centre, BC Cancer Agency
            cDNA preparation, sequencing and bioinformatics:
            Y Butterfield, R Kirkpatrick, J Asano, N Girn, R Guin, D Lee,
            S Lee, T Olson, P Pandoh, A Prahbu, D Smailus, L Spence, J Stott,
            S Taylor, G Yang, J Schein, S Jones and M Marra.
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .580
                     /organism="Salmo salar"
                     /mol_type="mRNA"
                     /strain="McConnell"
                     /db_xref="taxon:8030"
                     /clone_lib="mixed_tissue"
                     /note="Vector: pCMVsport6; Library Creator:  Research
                     Genetics ; Atlantic salmon tissue contributors: Carlo
                     Biagi, Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.),
                     Simon Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery
                     (Crofton, B.C.), Rachel Roper (University of Victoria)"
BASE COUNT      148 a    125 c    156 g    151 t
ORIGIN

  Query Match          9.5%;  Score 94.8;  DB 14;  Length 580;
  Best Local Similarity  63.7%;  Pred. No. 1.2e-11;
  Matches  144;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        196 TGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCT 255
              | || ||||   ||       ||   |||| || ||| | |||   ||| ||||||  |||
Db        308 TACTGGCTGGCTGGCTGACTGGCTGGCTGACTGGCTGGCAGGCTGGCTGGCTGGCTGGCT 367

Qy        256 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 315
                ||  || |||| |  ||| |||  ||  |||||  ||  ||| ||| ||||||   ||
Db        368 GACTAACTGGCTGACTGGCTGGCTGACTGACTGGCTGGCAGGCTGGCTGGCTGGCTGACT 427

Qy        316 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 375
              |||  ||  |||||  ||| |||  ||  ||| |   || ||| ||| ||||||   ||
Db        428 GGCTGACTGACTGGCTGGCTGGCTGACTAACTGACTGACTGGCTGGCTGGCTGGCTGACT 487

Qy        376 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGA 421
              ||| |||  ||||   ||| ||| |||  |||||  ||| |||||
Db        488 GGCTGGCTGACTGGTTGGCTGGCTGGCTGACTGGCTGGCTGGCTGA 533


RESULT 86
AZ364199/c
LOCUS       AZ364199                629 bp    DNA     linear   GSS 02-OCT-2000
DEFINITION  1M0110M16F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0110M16 F, genomic survey sequence.
ACCESSION   AZ364199
VERSION     AZ364199.1  GI:10477899
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 629)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0110  row: M  column: 16
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 629.
FEATURES             Location/Qualifiers
     source          1. .629
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0110M16"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
```

```
                    electrophoresis. Vector DNA was prepared from a derivative
                    of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                    inducible derivative of plasmid R1. The vector was ligated
                    with adaptors complementary to the insert adaptors and
                    purified. The sheared, adaptored mouse DNA was annealed to
                    adaptored vector DNA, and transformed into
                    chemically-competent E. coli XL10-Gold (Stratagene) cells
                    and selected for ampicillin resistance."
BASE COUNT      315 a    100 c    132 g     82 t
ORIGIN

  Query Match             9.5%;  Score 94.8;  DB 28;  Length 629;
  Best Local Similarity   69.4%;  Pred. No. 1.2e-11;
  Matches  129;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        234 TAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 293
              | | |||   | ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | |
Db        398 TGGTGCTTTGCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 339

Qy        294 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 353
              ||| ||| ||  ||  | ||||||||||||| ||| || ||||||||||||| ||| || ||
Db        338 CTTTCTTTCTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTG 279

Qy        354 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 413
              |||||||||| ||| || ||||||||||||  ||  | | ||| ||| ||  ||  | |
Db        278 CTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 219

Qy        414 CTTGCT 419
              ||| ||
Db        218 CTTTCT 213


RESULT 87
CB712482
LOCUS       CB712482                498 bp    mRNA    linear   EST 10-APR-2003
DEFINITION  AMGNNUC:NRDG1-00111-F7-A nrdg1 (10855) Rattus norvegicus cDNA clone
            nrdg1-00111-f7 5', mRNA sequence.
ACCESSION   CB712482
VERSION     CB712482.1  GI:29769630
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   Amgen EST Program.
  TITLE     Amgen Rat EST Program
  JOURNAL   Unpublished
COMMENT     Contact: Dan Fitzpatrick
            Amgen, Inc
            One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
            Tel: 805 447-4881
            Plate: 00111  row: f  column: 7.
FEATURES             Location/Qualifiers
     source          1. .498
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="nrdg1-00111-f7"
                     /tissue_type="Dorsal Root Ganglia"
                     /clone_lib="nrdg1 (10855)"
                     /note="Vector: pSPORT1; Site_1: Sal1; Site_2: Not1; rat
                     dorsal root ganglia"
BASE COUNT       21 a    222 c    119 g    136 t
ORIGIN

  Query Match             9.5%;  Score 94.6;  DB 14;  Length 498;
  Best Local Similarity   52.6%;  Pred. No. 1.2e-11;
  Matches  231;  Conservative    0;  Mismatches  204;  Indels    4;  Gaps    1;

Qy        240 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 299
              ||| | |||| ||||     | | | |||  ||||     ||| | |||| || |

Db         12 TGCCCTCTGGGCTGCCCCTGTACCCTCTGAGCTGCCCTTGTGCCCTCTGGGCTTCCCCTG 71

Qy        300 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 359
              ||  | |||| ||||    | ||| | |||| || |     ||| | |||  ||||
Db         72 TGTCCTCTGGGCTGCCCCCATGCCCTCTGGGCTTCCCCTGTGCCCTCTGAGCTGCCCCTG 131

Qy        360 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              ||| | |||  ||||     ||| | |||| ||||     ||| | ||   |    | |
Db        132 TGCCCTCTGAGCTGCCCCTGTGCCCTCTGGGCTGCCCCTGTGCCCTCTAAGC----TTCC 187

Qy        420 GAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGAT 479
                | | |    |||||||     | |     || | |  |   | ||   ||   ||
Db        188 CCTGTGCCCTCTGGGCTGCCCCTGTACCCTCCGGGCTGCCCCTGTACCCTCTGAGCTGCC 247

Qy        480 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 539
                | | |  |||||||  |   | | |  |||||||  |   |  | |  |||||||  |
Db        248 CTTGTGCCCTCTGGGCTTCCCCTGTGCCCTCTGGGCTTCCCCTGTGCCCTCTGGGCTTCC 307

Qy        540 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 599
                | | |  |||||||||   | | |  |||||||  |   |  | |  |||  |||||
Db        308 CCTGTGTCCTCTGGGCTGCCCCTGTGCCCTCTGGGCTTCCCCTGTGCCCTCTGAGCTGCC 367

Qy        600 AGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 659
                | | |  |||  |||||   | |    |||  |||||   | | |  |||  |||||
Db        368 CCTGTGCCCTCTGAGCTGCCCCTGTGCACTCTGAGCTGCCCCTGTGCCCTCTGAGCTGCC 427

Qy        660 AGTAGGGCTGCTGGGCTGC 678
                | | |  ||  |||
Db        428 CCTGTGCCCTCTAAGCTTC 446


RESULT 88
AZ282406
LOCUS       AZ282406                656 bp    DNA     linear   GSS 27-JUL-2000
DEFINITION  RPCI-23-138J12.TJ RPCI-23 Mus musculus genomic clone RPCI-23-138J12
            , genomic survey sequence.
ACCESSION   AZ282406
VERSION     AZ282406.1  GI:9524192
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 656)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished
COMMENT     Other_GSSs: RPCI-23-138J12.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 138  row: J  column: 12
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .656
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-138J12"
```

```
                     /sex="Female"
                     /lab_host="DH10B"
                     /clone_lib="RPCI-23"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).  "
BASE COUNT      205 a    163 c    282 g      6 t
ORIGIN

  Query Match             9.5%;  Score 94.6;  DB 28;  Length 656;
  Best Local Similarity   46.9%;  Pred. No. 1.3e-11;
  Matches  295;  Conservative    0;  Mismatches  334;  Indels    0;  Gaps    0;

Qy         86 CTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 145.
Db         25 CCGTCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGACAGACAGGCAGGCAGACAGGCAGG 84

Qy        146 ACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGA 205
Db         85 CAGGCAGGCAGGCAGGCAGGCAGACAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGACA 144

Qy        206 GGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCG 265
Db        145 CAGGCAGGCAGGCAGGCAGGCAGACAGGCAGGCATGTTGGCAGGCAGACAGGCAGGCAGG 204

Qy        266 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 325
Db        205 CAGGCAGACAGACAGGCAGGCAGGCAGACAGACAGGCAGGCAGGCAGGCAGGCAGGCAGG 264

Qy        326 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 385
Db        265 CAGACAGGCAGACAGGCAGGCAGGCAGGCAGACAGGCAGGCAGACAGGCAGGCAGGCAGG 324

Qy        386 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTG 445
Db        325 CAGACAGGCAGACAGGCAGGCAGACAGGCAGGCAGACAGGCAGACAGGCAGGCAGGCAGG 384

Qy        446 CTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGG 505
Db        385 CAGGCAGGCAGGCAGGCAGACAGGCAGACAGGCAGGCAGACAGGCAGGCAGGCAGGCAGA 444

Qy        506 GCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGG 565
Db        445 CAGGCAGGCAGGCAGGCAGACAGGCAGGCAGGCAGGCAGACAGGCAGGCAGGCAGGCAGG 504

Qy        566 GCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGG 625
Db        505 CAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGACAGGCAGGCAGGCAGG 564

Qy        626 GCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGG 685
Db        565 CAGACAGGCAGGCAGGCAGGCAGACAGGCAGACAGGCAGGCAGGCAGGCAGGCAGACAGG 624

Qy        686 GCTGCCTGCTGGCTTGCTTGCTTGCTTGC 714
Db        625 CAGGCAAGCAGACAGGCAGGCAGACAGGC 653


RESULT 89
BZ180072/c
LOCUS       BZ180072                621 bp    DNA     linear   GSS 11-OCT-2002
DEFINITION  CH230-302D17.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-302D17, genomic survey sequence.
ACCESSION   BZ180072
VERSION     BZ180072.1  GI:23824486
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 621)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 302  row: D  column: 17
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .621
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-302D17"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      223 a    189 c    124 g     85 t
ORIGIN

  Query Match             9.4%;  Score 94.4;  DB 28;  Length 621;
  Best Local Similarity   48.3%;  Pred. No. 1.5e-11;
  Matches  263;  Conservative    0;  Mismatches  281;  Indels    0;  Gaps    0;

Qy        225 GCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 284
Db        620 GCATGTTGGTGTACTTGCATGTTGGTGTACCTGCATGTTGGTGTACCTGCATGTTTGTGT 561

Qy        285 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 344
Db        560 ACATGCATGTTGGTGTACCTGCATGTTTGTGTGCCTGCATGTTGGTGTAGCTGCATGTTT 501

Qy        345 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 404
Db        500 GTGTACCTGCATGTTTGTGTGCTTGCATGTTTGTGTGCCTGCATGTTGGTGTACCTGCAT 441

Qy        405 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCT 464
Db        440 GTTGGTGTGCCTGCATGTTGGTGTATCTGCATGTTGGTGTACCTGCATGTTGGTGTACCT 381

Qy        465 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 524
Db        380 GCATGTTGGTGTACCTGCATGTTGGTGTACCTGCATGTTGGTGTACCTGCATGTTGGTGT 321

Qy        525 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 584
Db        320 ACCTGCATGTTTGTGTGCCTGCATGTTGGTGTAGCTGCATGTTTGTGTACCTGCATGTTT 261

Qy        585 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 644
Db        260 GTGTACCTGCATGTTTGTGTGCCTGCATGTTTGTGTACCTGCATGTTTGTGTGCCTGCAT 201

Qy        645 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTT 704
```

```
Db        200 GTTGGTGTACCTGCATGTTTGTGTGCCTGCATGCTTGTGTACCTGCATGTTTGTGTGACT 141

Qy        705 GCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACA 764

Db        140 GCATGTTTGGAACTGTCAGCCAGGGCCTGACAGGCCCACCATTGAGGACTCAGCTGAAAA 81

Qy        765 GAGT 768

Db         80 CAGT 77


RESULT 90
BB657463/c
LOCUS       BB657463               635 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB657463 RIKEN full-length enriched, 12 days embryo spinal ganglion
            Mus musculus cDNA clone D130098J21 5', mRNA sequence.
ACCESSION   BB657463
VERSION     BB657463.1  GI:16491289
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 635)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES    "       Location/Qualifiers
     source         1. .635
                    /organism="Mus musculus"
                    /mol_type="mRNA"
                    /db_xref="taxon:10090"
                    /clone="D130098J21"
                    /tissue_type="spinal ganglion"
                    /dev_stage="12 days embryo"
                    /lab_host="DH10B"
                    /clone_lib="RIKEN full-length enriched, 12 days embryo
                    spinal ganglion"
                    /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                    prepared and sequenced in Mouse Genome Encyclopedia
                    Project of Genome Exploration Research Group in Riken
                    Genomic Sciences Center and Genome Science Laboratory in
                    RIKEN. Division of Experimental Animal Research in Riken
                    contributed to prepare mouse tissues. 1st strand cDNA was
                    primed with a primer [5'
                    GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
                    prepared by using trehalose thermo-activated reverse
                    transcriptase and subsequently enriched for full-length by
                    cap-trapper. Second strand cDNA was prepared with the
                    primer adapter of sequence [5'
                    GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                    was cleaved with BamHI and XhoI. Vector: a modified
                    pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT      215 a    192 c    134 g     93 t      1 others
ORIGIN

  Query Match             9.4%;  Score 94.4;  DB 10;  Length 635;
  Best Local Similarity   62.9%;  Pred. No. 1.5e-11;
  Matches  146;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        184 TGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCT 243

Db        376 TGGTGGCTGCTGCTCTCCCTGTTGTTCCTGTTGCCACTTTGGATTGCTGGATTGCTGGAT 317

Qy        244 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 303

Db        316 TGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCT 257

Qy        304 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 363

Db        256 GGATTGCTGGATTGCTGGATTGCTGGATTGCTGGGTTGCTGGGTTGCTGGGTTGCTGGAT 197

Qy        364 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 415

Db        196 TGCTGGATTGCTGGGTTGCTGGATTGCTGGATTGCTGGATTGCTGGGTTGCT 145


RESULT 91
AK084134/c
LOCUS       AK084134               1673 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 12 days embryo spinal ganglion cDNA, RIKEN full-length
            enriched library, clone:D130098J21 product:hypothetical protein,
            full insert sequence.
ACCESSION   AK084134
VERSION     AK084134.1  GI:26350970
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
```

```
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 1673)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1...1673
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:D130098J21"
                     /db_xref="taxon:10090"
                     /clone="D130098J21"
                     /tissue_type="spinal ganglion"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="12 days embryo"
     CDS             <1. .472
                     /note="unnamed protein product; hypothetical protein
                     (evidence: NCBI CDS Predictor,Truncated-Longest-ORF)
                     putative"
                     /codon_start=2
                     /protein_id="BAC39122.1"
                     /db_xref="GI:26350971"
                     /translation="TTGGTVTEDREESREGQLQRKVSGESLPGVRDHAGSTITGGKGR
                     LQCRQSNPAIQQSSNPATQQSSNPATQQPSNPAIQQSSNPAIQQSSNPAIQQSSNPAI
                     QQSSNPAIQQSKVATGTTGRAAATTDPLGALAFIPSPGSPDFQTICEWQEPHPS"
BASE COUNT      511 a    419 c    434 g    309 t
ORIGIN

  Query Match             9.4%;  Score 94.4;  DB 11;  Length 1673;
  Best Local Similarity   62.9%;  Pred. No. 1.9e-11;
  Matches  146;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        184 TGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCT 243
              || | ||||     ||  ||| |   | || |      ||| ||||| |   |   | |
Db        380 TGGTGGCTGCTGCTCTCCCTGTTGTTCCTGTTGCCACTTTGGATTGCTGGATTGCTGGAT 321

Qy        244 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 303
              |||||  |||| | ||||| | | ||  | ||||| | | ||||| |||| | |||||
Db        320 TGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCT 261

Qy        304 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 363
              | | ||  | ||||| | | |||||  |||| | ||||| | | ||  | ||||| | |
Db        260 GGATTGCTGGATTGCTGGATTGCTGGATTGCTGGGTTGCTGGGTTGCTGGGTTGCTGGAT 201

Qy        364 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 415
              |||||  |||| | ||||| | | ||  | ||||| | | |||||  ||||
Db        200 TGCTGGATTGCTGGGTTGCTGGATTGCTGGATTGCTGGATTGCTGGGTTGCT 149


RESULT 92
CNS024NL
LOCUS       CNS024NL                1152 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            235E04 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL181002
VERSION     AL181002.1  GI:7819059
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
```

```
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1152)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1152
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="235E04"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG235BC02SP1-end :
                     PUC-Ori"
BASE COUNT      306 a     97 c    347 g    318 t     84 others
ORIGIN

  Query Match             9.4%;  Score 93.8;  DB 29;  Length 1152;
  Best Local Similarity   54.2%;  Pred. No. 2.4e-11;
  Matches  173;  Conservative    7;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        187 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGC 246
              | ||||||  ||| | ||   ||   || ||   | ||| ||| |    ||   | | ||
Db        754 TGGCTGGCTGGCTGGTTGGTTGGTTGGTTGGTTGGTTTGTTGACTGGTTGGTTGGC 813

Qy        247 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 306
              || |  ||| ||| | | | |||   | | | |  || | |||   ||| ||| ||| ||
Db        814 TGACTGGCTGGCTGGTTGGTTGGTTGGTTGGTTGACTGGTTGGTTGGCTGGCTGGCTGGC 873

Qy        307 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 366
              ||||  | | | | | | | |||   | ||| |  :| | |||   | | ||| ||| ||
Db        874 TGGCTGGTTGGTTGGTTKGTTGGTTGGTTTGTTGAYTGGTTGGTTGGTTGGCTGGCTGGC 933

Qy        367 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 426
              |||  | | | | | | | | || |  ||| ||| |:| | |||  || |  |||   ||
Db        934 TGGTTGGTTGGTTGGTTTGTTGACTGGCTGGCTGGYTGGTTGGTTTGTTGACTGGTTGGT 993

Qy        427 CAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGG 486
              :  ||   : |:| ||| | |  ||   |   | ||: ||| : | ||| | |   ||
Db        994 TGKCTGGCTGKTTKGTTTGTTGACTGGTTTGGTTGGCTKGCTGGYTGGTTGGTTGGTTGG 1053

Qy        487 CTGCTGGGCTGCTAGTAGG 505
              ||| | |   || || ||
Db       1054 CTGGTTGTTGACTGGTTGG 1072


RESULT 93
BH731712/c
LOCUS       BH731712               490 bp    DNA     linear   GSS 20-FEB-2002
DEFINITION  BOMAD94TF BO_2_3_KB Brassica oleracea genomic clone BOMAD94,
            genomic survey sequence.
ACCESSION   BH731712
VERSION     BH731712.1  GI:18837107
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOMAD94TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .490
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOMAD94"
                     /clone_lib="BO_2_3_KB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      100 a    167 c    174 g     49 t
ORIGIN

  Query Match             9.3%;  Score 93.2;  DB 28;  Length 490;
  Best Local Similarity   50.0%;  Pred. No. 2.6e-11;
  Matches  233;  Conservative    0;  Mismatches  233;  Indels    0;  Gaps    0;

Qy        245 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 304
              ||  ||| | | |||   | ||  ||| | | | |   | ||  ||| | | |||   |
Db        489 GCGTGCCCGTTCGCTACGTAGCGTGCCCGGTCGTTACGTAGCGTGCCCGGTCGCTACGTA 430

Qy        305 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 364
              ||  ||| | | |||   | ||  ||| | | |||   | ||  ||| | | |||   |
Db        429 GCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTA 370

Qy        365 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 424
              ||  ||| | | |||   | ||  ||| | | |||   | ||  ||| | | |||   |
Db        369 GCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTA 310

Qy        425 GACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAG 484
              |   |   ||  |||   | ||      |   |       || ||   | | | ||
Db        309 GCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTA 250

Qy        485 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 544
              |  |||  ||  ||||    |  |||  ||  ||||    |  |||  ||  ||||
Db        249 GCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTA 190

Qy        545 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 604
              |  |||  ||  ||||    |  | |  ||  ||||    |  |||  ||  ||||
Db        189 GCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTA 130

Qy        605 GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 664
              |  | |  ||  ||||    |  |||  ||  ||||    |  |||  ||  ||||
Db        129 GCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCTTGCCCGGTCGCTACGTA 70

Qy        665 GGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGC 710
              |  |||  ||  ||||    |  |||| | | |||   | ||  |||
Db         69 GCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCCTGC 24


RESULT 94
CA063770/c
LOCUS       CA063770               737 bp    mRNA    linear   EST 04-MAR-2003
DEFINITION  ssalrgb530042 mixed_tissue Salmo salar cDNA, mRNA sequence.
ACCESSION   CA063770
VERSION     CA063770.1  GI:24394013
KEYWORDS    EST.
SOURCE      Salmo salar (Atlantic salmon)
  ORGANISM  Salmo salar
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
REFERENCE   1  (bases 1 to 737)
  AUTHORS   GRASP Consortium, Davidson,W.S., Koop,B.F. and
            http://web.uvic.ca/cbr/grasp.
```

```
  TITLE     A survey of Salmo salar transcripts from high complexity cDNA
            libraries
  JOURNAL   Unpublished
COMMENT     Contact: Koop BF
            Centre for Biomedical Research
            University of Victoria
            PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
            Tel: 250 472 4067
            Fax: 250 472 4075
            Email: bkoop@uvic.ca
            Genome Sciences Centre, BC Cancer Agency
            cDNA preparation, sequencing and bioinformatics:
            Y Butterfield, R Kirkpatrick, J Asano, N Girn, R Guin, D Lee,
            S Lee, T Olson, P Pandoh, A Prahbu, D Smailus, L Spence, J Stott,
            S Taylor, G Yang, J Schein, S Jones and M Marra.
FEATURES             Location/Qualifiers
     source          1. .737
                     /organism="Salmo salar"
                     /mol_type="mRNA"
                     /strain="McConnell"
                     /db_xref="taxon:8030"
                     /clone_lib="mixed_tissue"
                     /note="Vector: pCMVsport6; Library Creator:  Research
                     Genetics ; Atlantic salmon tissue contributors: Carlo
                     Biagi, Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.),
                     Simon Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery
                     (Crofton, B.C.), Rachel Roper (University of Victoria)"
BASE COUNT      274 a    185 c    163 g    115 t
ORIGIN

  Query Match             9.3%;  Score 93.2;  DB 14;  Length 737;
  Best Local Similarity   48.6%;  Pred. No. 2.9e-11;
  Matches  286;  Conservative    0;  Mismatches  298;  Indels    4;  Gaps    1;

Qy        146 ACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGA 205
              |||||  | |||| ||       | |    | |||  ||   |||    ||   |
Db        737 ACAGTCTGTCTGCCTGTCTGACTGCTATGGTGTCTGTCTGTCTGCTATGGTGACAGTCTG 678

Qy        206 GGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCG 265
                | | ||  | |||   |  |||||    | | |   |   | |||  ||  ||   |
Db        677 TCTGCCTGTCTGTCTGTCTGACTGCTATGGTGACAGTCTGTCTGTCTGTCTGTCTGACTG 618

Qy        266 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 325
              ||    ||   |  ||    |  | ||| ||  ||   |   | |||  ||  ||   |
Db        617 CTATGGTGACAGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGACTG 558

Qy        326 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 385
              |     ||   |  ||   ||  | |||  ||||     |   | |||  ||||     |
Db        557 CAATGGTGACAGTCTGTCTGCCTGTCTGACTGCTATGGTGTCTGTCTGTCTGCTATGGTG 498

Qy        386 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTG 445
                | |||  ||| ||   |   | |||  ||||  || ||||||    |||  ||  ||
Db        497 ACAGTCTGTCTGCCTGTCTGTCTGTCTGACTGCTATGGTGACAGT----CTGTCTGTCTG 442

Qy        446 CTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGG 505
              ||| || | | ||  | ||  ||| |    ||| |||  | ||||  | |||   ||  |
Db        441 CTATGGTGTCTGTCTGTCTAACTGCTATGGTGACAGTCTGTCTGCCTGTCTGTCTGTCTG 382

Qy        506 GCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGG 565
              |||||  | ||  |||  | |||   | |||   |   |       | | |   ||  |
Db        381 ACTGCTATGGTGACAGTCTGTCTGTCTGTCTGTCTGACTGCTATGGTGACAGTCTGTCTG 322

Qy        566 GCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGG 625
              ||    | |||| |    | | |   | ||||  ||  | || ||  | ||   ||  |
Db        321 TCTGTATGACTGCAATGGTGACAGTCTGTCTGCCTGTCTGACTGCTATGGTGTCTGTCTG 262

Qy        626 GCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGG 685
              |||||  | ||  |||  | ||||  | |||   ||  | ||||   | ||  |||  |
Db        261 TCTGCTATGGTGACAGTCTGTCTGCCTGTCTGTCTGTCTGACTGCAATGGTGACAGTCTG 202

Qy        686 GCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTG 733
              |||||||   |  ||||    ||  ||   ||  |      || |||
Db        201 TCTGCCTGTCTGTCTGCTATGGTGTCTGTCTGTCTGCTATGGTGTCTG 154
```

```
RESULT 95
CNS04CCM
LOCUS       CNS04CCM                1057 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            099C18 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL284287
VERSION     AL284287.1  GI:8022668
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1057)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1057
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="099C18"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG099BB09LP1-end : T7"
BASE COUNT      186 a    307 c    286 g    265 t     13 others
ORIGIN

  Query Match             9.3%;  Score 93.2;  DB 29;  Length 1057;
  Best Local Similarity   56.0%;  Pred. No. 3.2e-11;
  Matches  167;  Conservative    4;  Mismatches  127;  Indels    0;  Gaps    0;

Qy        210 ACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 269
              |||||| | |||   || |||  ||  | ||||  |   |||||  ||  ||    | |
Db        541 ACAGTAAGTCTGTCTGCCTGCCTGTCTGCCTGCCTGTCTGCCTGTCTGTCTGTCTACCTG 600

Qy        270 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 329
              ||||| ||| ||   ||  ||||   ||  |||  |   |||||| |   ||   ||  |
Db        601 CCTGCCTGCCTGTCTGCCTGCCTATCTGTCTGCCTGTCTGCCTGCCTATCTGTCTGCCTG 660

Qy        330 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 389
              ||||  ||| |||  ||  | |||| ||| ||   |   |||||  ||| |||  |   |
Db        661 TCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTGTCTGCCTGTCTGCCTGCCTGTCTG 720

Qy        390 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAA 449
```

```
Db        721 CCTGTCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGTCTGTCTGTCTGCCTGTCTGCCTG 780

Qy        450 GGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGC 507

Db        781 TCTGTCTGTCTGCCWGCCTRWMTGCMTGTCTGTCTGTCTGCCTGCCWGTCTGTCTGTC 838


RESULT 96
FR0038988/c
LOCUS       FR0038988               461 bp    DNA     linear   GSS 22-OCT-1999
DEFINITION  Fugu rubripes GSS sequence, clone 089D07bA10, genomic survey
            sequence.
ACCESSION   AL126486
VERSION     AL126486.1  GI:6108101
KEYWORDS    GSS; genome survey sequence.
SOURCE      Takifugu rubripes
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K.,
            Umrania,Y., Williams,G. and Brenner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
            Centre, Hinxton, Cambridge, CB10 1SB. UK Email:
            biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Takifugu rubripes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:31033"
                     /clone="089D07bA10"
                     /clone_lib="cosmid 089D07"
BASE COUNT      112 a    170 c     80 g     70 t     29 others
ORIGIN

  Query Match             9.3%;  Score 92.8;  DB 29;  Length 461;
  Best Local Similarity   54.2%;  Pred. No. 3.2e-11;
  Matches  166;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy        234 TAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 293

Db        426 TAGGTACCCTGGCTGGNTGAGTGGCTGGNTAGGTACNTCGCTAGGTGAGTGNCTGGCTAG 367

Qy        294 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 353

Db        366 GTAGCTGGCTGGCTGNNTGGGTGAGTGACTGGCTAGGNAGCTGGCTAGGTGGCTGNCTAG 307

Qy        354 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 413

Db        306 GTGNNTAGGTAGCTGNCTGGGTGAGTGGCTGGCTGGGTGAGTGGNTAGNTAGCTGGCTAG 247

Qy        414 CTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGG 473

Db        246 NTAGCTGGCTAGNTAGGTGGCTAGGTGGCTGGCTAGGTGGCTAGGTTGCTGGCNAGGTGG 187

Qy        474 CTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 533

Db        186 CTAGGTAGGTGGCTAGGTAGGTGGCTAGGTGGCTGGCTAGGTGGCTAGGTAGGTGGCTAG 127

Qy        534 GCTGCT 539

Db        126 GTNNCT 121


RESULT 97
AL562271
LOCUS       AL562271               1200 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL562271 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DC003YH19 3-PRIME, mRNA sequence.
ACCESSION   AL562271
VERSION     AL562271.2  GI:31286286
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1200)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12910526.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6688.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DC003CD10NP1&cluster=6688.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DC003CD10NP1.
FEATURES             Location/Qualifiers
     source          1. .1200
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DC003YH19"
                     /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT       70 a    371 c    442 g    129 t    188 others
ORIGIN

  Query Match             9.3%;  Score 92.6;  DB 9;  Length 1200;
  Best Local Similarity   40.5%;  Pred. No. 4.5e-11;
  Matches  303;  Conservative   59;  Mismatches  382;  Indels    5;  Gaps    1;

Qy         12 CGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGC 71

Db         38 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 97

Qy         72 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 131

Db         98 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 157

Qy        132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191

Db        158 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 217

Qy        192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251

Db        218 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGTGTGCGCGCGCGCGCGCG 277

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311

Db        278 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGYGCGCGCGCG 337

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371

Db        338 YGCGCGCGCGYGCGCGCGCGCGCGYGYGYGCGCGCGCGCGCGCGCGCGCGCGCGCGYG 397
```

```
Qy      372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTA 431
            ||  || :|: |||  |:  ||  || ||: :||  ||  ||  ||    | |   |:
Db      398 CGCGCGCGYGMGCGCGYGYGCGCGCGCGTGYGYGCGCGCGCGCGCGCGCGYGYGYGCGYG 457

Qy      432 GGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCT 491
            |   ||  || :|:    | |:    :: : ||: ::   |: :| : ||:   |:|
Db      458 MGYGCGCGCGCGYGHGTGTGYGHRYACRYAYGTGYGYRTGTGYAYGTKTGTMTMHCYGTG 517

Qy      492 GGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCT 551
            |  ||: ::|  :  ||:  |  ||: :|:    |  |   ||: :  |  | :||
Db      518 TGTGTGYGMSTGWSTGTGYGTGTGTGHGMGYGTTTTTTTTTTHCTKTKTTTKTGTSTGTG 577

Qy      552 GGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCT 611
            |  ||   |   | : | |   |   |  ||: |::  |   :|  |:  |
Db      578 TGTGTGTGTGCGCGCGYGCGCGCGCGTGYGCGGGKGTVYGTGTGCSCGYGYGCGCGCGTG 637

Qy      612 GGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCT 671
            |  ||:  ||  |  ::   |  ||   ||  |  ||   |  :|| |:  :  :|:
Db      638 TGTGTGYGYGTGTGTGYSTGTGTGTGTGTGTGYGTGTGTGTGYGCGYGYGYRTGYGYR 697

Qy      672 GGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCC 731
            |  ||   ||   |     ||   |: ||  ||  ||  |||  ||| |  :|: ||
Db      698 TGTGTGTGTGTGTAG-----TGTGCGYGTGTGTGTGTGTGTGCGCGTGTGTGYGYGTGTG 752

Qy      732 TGCTACTAGGGCTGCTGTGCAGCTGGGAG 760
            |    |:|: |:   ||||:   || | |
Db      753 TATCYCYASYGYGYGTGTGYGTGTGTGTG 781
```

RESULT 98
BB651343

```
LOCUS       BB651343                622 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB651343 RIKEN full-length enriched, 0 day neonate cerebellum Mus
            musculus cDNA clone C230076M22 5', mRNA sequence.
ACCESSION   BB651343
VERSION     BB651343.1  GI:16485596
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 622)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .622
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="C230076M22"
                     /tissue_type="cerebellum"
                     /dev_stage="0 day neonate"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, 0 day neonate
                     cerebellum"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 479.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
BASE COUNT       73 a    233 c     32 g    278 t      6 others
ORIGIN

  Query Match             9.2%;  Score 92.2;  DB 10;  Length 622;
  Best Local Similarity   66.2%;  Pred. No. 4.7e-11;
  Matches  133;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy      219 CTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 278
            || ||  |||  |  ||   ||||| || || |||||||||| ||  ||  | | ||| |
Db       59 CTCCTCCCTTCTTTATATCTCTGCTTTCTTGCTTGCTTTCTTTCTTTCTTTCTTTC 118

Qy      279 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 338
            || ||  ||  | | |||||||||| ||| || |||||||||||| ||| || |||||||
Db      119 TTTCTTTCTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 178

Qy      339 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 398
            |||||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| |
Db      179 TTGCTTTCTTTCTTTCTTTCTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCC 238

Qy      399 TTGCTCGCTGGCCTGCTTGCT 419
            || ||  |   | | ||  ||
Db      239 TTCCTTCCCTCCTTTCTCTCT 259
```

RESULT 99
BH325278

```
LOCUS       BH325278                661 bp    DNA     linear   GSS 03-DEC-2001
DEFINITION  CH230-45J13.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
            CH230-45J13, genomic survey sequence.
ACCESSION   BH325278
VERSION     BH325278.1  GI:17255992
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
```

```
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 661)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 EcoRI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-45J13.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 45  row: J  column: 13
            Seq primer: SP6
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .661
                      /organism="Rattus norvegicus"
                      /mol_type="genomic DNA"
                      /strain="BN/SsNHsd/MCW"
                      /db_xref="taxon:10116"
                      /clone="CH230-45J13"
                      /sex="Female"
                      /cell_type="Brain"
                      /clone_lib="CHORI-230 Segment 1"
                      /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                      CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                      Pieter de Jong"
BASE COUNT      172 a    178 c    138 g    173 t
ORIGIN

  Query Match             9.2%;  Score 92.2;  DB 28;  Length 661;
  Best Local Similarity   73.3%;  Pred. No. 4.8e-11;
  Matches  118;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy        232 AGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 291
              | ||  | |  | ||| ||||||| |   ||| ||| |  |||||||||||| ||| ||
Db        501 AATAATGGTATTAGCTAGCCTGCTGGTGGGCTTGCTAGGTTGCTTGCTTGCTTGCTTGCT 560

Qy        292 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 351
              |||||||||||| |||  | | ||||||| || |||  | ||||| ||||||  || ||
Db        561 TGCTTGCTTGCTTGCTTTCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTTGCT 620

Qy        352 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 392
              | |||||||||| ||| || |||| ||||||  || || |
Db        621 TCCTTGCTTGCTTGCTTGCTTGCTTCCTTGCTTCCTTGCTT 661


RESULT 100
AZ501999/c
LOCUS       AZ501999                760 bp    DNA     linear   GSS 05-OCT-2000
DEFINITION  1M0341N01F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0341N01 F, genomic survey sequence.
ACCESSION   AZ501999
VERSION     AZ501999.1  GI:10683315
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 760)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0341  row: N  column: 01
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 760.
FEATURES              Location/Qualifiers
     source           1. .760
                      /organism="Mus musculus"
                      /mol_type="genomic DNA"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC1M0341N01"
                      /sex="Male"
                      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                      /clone_lib="Mouse 10kb plasmid UUGC1M library"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (male) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
                      adaptored DNA was purified and size-selected for a 9.5 to
                      10.5 kb range using preparative agarose gel
                      electrophoresis. Vector DNA was prepared from a derivative
                      of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                      inducible derivative of plasmid R1. The vector was ligated
                      with adaptors complementary to the insert adaptors and
                      purified. The sheared, adaptored mouse DNA was annealed to
                      adaptored vector DNA, and transformed into
                      chemically-competent E. coli XL10-Gold (Stratagene) cells
                      and selected for ampicillin resistance."
BASE COUNT      239 a    170 c    189 g    162 t
ORIGIN

  Query Match             9.2%;  Score 92.2;  DB 28;  Length 760;
  Best Local Similarity   63.7%;  Pred. No. 5e-11;
  Matches  156;  Conservative    0;  Mismatches   88;  Indels    1;  Gaps    1;

Qy        175 AGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGT 234
              | ||||||||||  ||  |||||| || || |  | | ||  | ||||  || |||  |
Db        280 ATGCCTGCTTGCCTGCCTGCCTGCCTGTTG-GCTGTCTGTTTGCCTGCCTGCCTGCCTGC 222

Qy        235 AGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 294
                | |||    |   |  ||| ||| |||  ||  ||||| |||| |||  ||  |||||
Db        221 CTGCCTGGCTGTCTGTTTGCCTGCCTGCCTGCCTGCCTGTTTGCCTGCCTGCCTGCCTGG 162

Qy        295 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 354
              ||  || |  |  ||||||||||| |||  |   | ||| |||| |||  ||  ||||||
Db        161 CTGTCTGTTTACCTGCCTGCTTGCCTGCCTGGCTGTCTGTTTGCCTGCCTGCCTGCCTGC 102

Qy        355 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 414
              ||  ||   | |  ||||| |||||||  ||  | |||  || ||||  ||  ||||||
Db        101 CTGGCTGTCTGTTTACCTGCCTGCTTGCCTGCCTGGCTGTCTGTTTGCCTGCCTGCCTGC 42

Qy        415 TTGCT 419
```

```
                 ||||
Db          41 CTGCT 37

Search completed: August 22, 2003, 00:38:54
Job time : 2105.46 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 21, 2003, 18:43:32 ; Search time 3718.64 Seconds
(without alignments)
11001.231 Million cell updates/sec

Title: US-09-831-000-1_COPY_1_1000
Perfect score: 1000
Sequence: 1 gatcgggaaaacgcgagggg.....gtgcacggccaagttacaag 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1000	100.0	133719	14	AF083501	AF083501 Macaca mu
c	2	286.6	28.7	148106	2	AC116487	AC116487 Mus muscu
	3	280.6	28.1	265401	2	AC099177	AC099177 Rattus no
	4	272.6	27.3	165180	2	AC119239	AC119239 Mus muscu
c	5	240.8	24.1	218921	2	AC097912	AC097912 Rattus no
	6	239.2	23.9	130733	14	AF210726	AF210726 Macaca mu
c	7	234.4	23.4	229535	2	AC130110	AC130110 Rattus no
c	8	226.2	22.6	245853	2	AC131862	AC131862 Rattus no
	9	226.2	22.6	305166	2	AC106141	AC106141 Rattus no
c	10	225.2	22.5	208050	2	AC137146	AC137146 Mus muscu
c	11	225.2	22.5	228919	2	AC112675	AC112675 Mus muscu
	12	218.6	21.9	220464	2	AC115409	AC115409 Rattus no
	13	218.2	21.8	128953	9	AC107913	AC107913 Homo sapi
	14	213	21.3	238215	2	AC141883	AC141883 Mus muscu
	15	202.6	20.3	219713	2	AC132225	AC132225 Mus muscu
	16	202.6	20.3	229010	2	AC073686	AC073686 Mus muscu
	17	202	20.2	276954	2	AC098079	AC098079 Rattus no
	18	201.2	20.1	271190	2	AC117991	AC117991 Mus muscu
	19	195.4	19.5	224055	2	AC096116	AC096116 Rattus no
c	20	194.4	19.4	183863	2	AC110455	AC110455 Rattus no
	21	192.8	19.3	27213	9	AP000291	AP000291 Homo sapi
	22	192.8	19.3	100000	9	AP000043	AP000043 Homo sapi
	23	192.8	19.3	100000	9	AP000111	AP000111 Homo sapi
	24	192.8	19.3	100000	9	AP000187	AP000187 Homo sapi
	25	192.8	19.3	340000	9	AP001716	AP001716 Homo sapi
	26	192.2	19.2	234939	2	AC094809	AC094809 Rattus no
c	27	191.6	19.2	235855	2	AC094484	AC094484 Rattus no
c	28	191.4	19.1	189096	5	AL935202	AL935202 Zebrafish
	29	190.6	19.1	223156	2	AC112928	AC112928 Mus muscu
	30	189.4	18.9	209013	10	AL844846	AL844846 Mouse DNA
c	31	189.2	18.9	208740	2	AC114591	AC114591 Mus muscu
	32	187.6	18.8	230500	2	AC097040	AC097040 Rattus no
c	33	187.6	18.8	258132	2	AC142075	AC142075 Rattus no
c	34	186.2	18.6	203152	2	AC131371	AC131371 Rattus no
c	35	186.2	18.6	270245	2	AC095180	AC095180 Rattus no
c	36	184.4	18.4	133661	14	U93872	U93872 Kaposi's sa
	37	182.8	18.3	28199	5	AF013613	AF013613 Fugu rubr
	38	182.8	18.3	28199	6	AX652127	AX652127 Sequence
	39	182.4	18.2	110000	2	AC109525_0	AC109525 Rattus no
c	40	182.4	18.2	237019	2	AC094443	AC094443 Rattus no
	41	179.8	18.0	201258	2	AC141889	AC141889 Mus muscu
	42	178.6	17.9	235808	2	AC091372	AC091372 Rattus no
c	43	178.6	17.9	305226	2	AC107437	AC107437 Rattus no
c	44	177.6	17.8	271233	2	AC096597	AC096597 Rattus no
	45	177.4	17.7	599	9	AF388526	AF388526 Homo sapi
	46	177.4	17.7	88716	2	AC098644	AC098644 Takifugu
	47	177.4	17.7	275443	2	AC096407	AC096407 Rattus no
c	48	176.4	17.6	246968	2	AC113967	AC113967 Mus muscu
c	49	174.8	17.5	201258	2	AC141889	AC141889 Mus muscu
	50	173.4	17.3	242482	2	AC115149	AC115149 Rattus no
c	51	171	17.1	245258	2	AC113840	AC113840 Rattus no
	52	170.6	17.1	200141	10	AL845305	AL845305 Mouse DNA
	53	170.6	17.1	206805	2	AC118832	AC118832 Rattus no
	54	170.6	17.1	256110	2	AC116951	AC116951 Mus muscu
c	55	168	16.8	201016	2	AC120242	AC120242 Rattus no
	56	167.4	16.7	197153	2	AC141966	AC141966 Rattus no
c	57	167.4	16.7	228645	2	AC094273	AC094273 Rattus no
c	58	167.4	16.7	257675	2	AC094136	AC094136 Rattus no
c	59	166.8	16.7	197891	2	AC109837	AC109837 Rattus no
	60	166.4	16.6	205515	2	AC118905	AC118905 Rattus no
	61	166.4	16.6	272652	2	AC109709	AC109709 Rattus no
	62	165.2	16.5	234474	2	AC103472	AC103472 Rattus no
c	63	165.2	16.5	235837	2	AC112613	AC112613 Rattus no
c	64	165.2	16.5	256486	2	AC106350	AC106350 Rattus no
	65	165	16.5	152054	10	AC122009	AC122009 Mus muscu

	66	164.8	16.5 210964	2	AC113245
	67	164	16.4 196519	2	AC128183
	68	163.8	16.4 252801	2	AC095211
c	69	163.6	16.4 160805	2	AC116801
	70	162.6	16.3 191953	2	AC119267
	71	160.8	16.1 243716	2	AL833804
	72	160.6	16.1 176782	10	AC124492
c	73	160.2	16.0 231321	2	AC098121
c	74	160.2	16.0 238112	2	AC106600
c	75	160	16.0 110000	10	AE014176_0
c	76	160	16.0 189576	2	AC114585
c	77	159.8	16.0 110000	2	AC108277_1
	78	159.8	16.0 238255	2	AC132758
	79	159.8	16.0 240200	2	AC116206
	80	159.8	16.0 242974	2	AC103570
c	81	159.8	16.0 244301	2	AC116072
c	82	159.2	15.9 251656	2	AC106240
c	83	158.8	15.9 208994	2	AC122561
	84	158.8	15.9 217837	10	AL772377
	85	158.2	15.8 179192	2	AC044848
c	86	158.2	15.8 220050	10	AL583887
	87	158.2	15.8 236344	10	AL583889
	88	157.6	15.8 148066	10	AL606528
	89	157.2	15.7 189236	2	AL645694
c	90	157.2	15.7 217604	2	AC118179
	91	157.2	15.7 223892	2	AC110123
c	92	157.2	15.7 262289	2	AC117951
	93	156.8	15.7 176961	5	AL928917
c	94	155.8	15.6 140898	2	AC142424
	95	154.4	15.4 174032	2	AC128150
c	96	154.4	15.4 253076	2	AC111479
	97	154.2	15.4 283593	2	AC131433
	98	153.8	15.4 223379	2	AC106435
	99	153.8	15.4 265707	2	AC111671
	100	152.8	15.3 238046	2	AC095003
c	101	152.6	15.3 157624	10	AL669980
c	102	151.6	15.2 172445	2	AC128287
c	103	151.4	15.1 96396	2	AC133490
c	104	151.4	15.1 226287	2	AC123086
c	105	150.2	15.0 291205	2	AC103125
	106	149.8	15.0 164138	2	AC109432
	107	149.6	15.0 238488	2	AC103134
c	108	148.6	14.9 165118	2	AC119053
	109	147.6	14.8 240800	2	AC095964
	110	147.6	14.8 274697	2	AC098240
	111	147.4	14.7 189272	2	AC119503
c	112	147.2	14.7 209093	2	AC109607
c	113	147	14.7 251625	2	AC095726
	114	146.6	14.7 278905	2	AC126641
	115	146.4	14.6 165608	2	AC087842
	116	146.4	14.6 229887	2	AC073297
	117	146.2	14.6 218800	2	AC094410
c	118	145.8	14.6 231764	2	AC119470
c	119	145.8	14.6 280417	2	AC095158
	120	145.4	14.5 212491	10	AL929228
	121	145.4	14.5 236086	2	AC107243
c	122	145.4	14.5 246070	2	AC135737
c	123	145	14.5 201290	10	AL731682
	124	144.8	14.5 98228	2	AC135405
c	125	144.8	14.5 133719	14	AF083501
c	126	144.6	14.5 190175	2	AC137330
	127	144.6	14.5 247754	2	AC125950
	128	144.4	14.4 230283	2	AC119802
	129	144.2	14.4 267991	2	AC112578
	130	143.6	14.4 187446	2	AC124374
	131	143.6	14.4 225027	2	AC131725
c	132	143.2	14.3 213248	2	BX465859
	133	143	14.3 187995	10	AC127254
c	134	142.6	14.3 253951	2	AC128478
c	135	142.4	14.2 288551	2	BX322534
	136	142	14.2 76861	2	AC087121
c	137	142	14.2 242619	2	AC115464
c	138	142	14.2 264297	2	AC094040

AC113245 Mus muscu
AC128183 Rattus no
AC095211 Rattus no
AC116801 Mus muscu
AC119267 Mus muscu
AL833804 Mus muscu
AC124492 Mus muscu
AC098121 Rattus no
AC106600 Rattus no
AE014176 Mus muscu
AC114585 Mus muscu
Continuation (2 of
AC132758 Rattus no
AC116206 Rattus no
AC103570 Rattus no
AC116072 Rattus no
AC106240 Rattus no
AC122561 Mus muscu
AL772377 Mouse DNA
AC044848 Mus muscu
AL583887 Mouse DNA
AL583889 Mouse DNA
AL606528 Mouse DNA
AL645694 Mus muscu
AC118179 Rattus no
AC110123 Rattus no
AC117951 Rattus no
AL928917 Zebrafish
AC142424 Rattus no
AC128150 Rattus no
AC111479 Rattus no
AC131433 Rattus no
AC106435 Rattus no
AC111671 Rattus no
AC095003 Rattus no
AL669980 Mouse DNA
AC128287 Rattus no
AC133490 Rattus no
AC123086 Rattus no
AC103125 Rattus no
AC109432 Rattus no
AC103134 Rattus no
AC119053 Danio rer
AC095964 Rattus no
AC098240 Rattus no
AC119503 Rattus no
AC109607 Mus muscu
AC095726 Rattus no
AC126641 Rattus no
AC087842 Rattus no
AC073297 Mus muscu
AC094410 Rattus no
AC119470 Rattus no
AC095158 Rattus no
AL929228 Mouse DNA
AC107243 Rattus no
AC135737 Mus muscu
AL731682 Mouse DNA
AC135405 Rattus no
AF083501 Macaca mu
AC137330 Rattus no
AC125950 Rattus no
AC119802 Mus muscu
AC112578 Rattus no
AC124374 Mus muscu
AC131725 Mus muscu
BX465859 Danio rer
AC127254 Mus muscu
AC128478 Rattus no
BX322534 Homo sapi
AC087121 Mus muscu
AC115464 Rattus no
AC094040 Rattus no

	139	141.8	14.2 110000	2	AC115960_0
	140	141.8	14.2 208254	2	AC119002
	141	141.8	14.2 210320	2	AC140324
c	142	141.8	14.2 241649	2	AC120675
c	143	141.4	14.1 243164	2	AC132173
	144	141.4	14.1 253902	2	AC095908
	145	141.2	14.1 190073	2	AC138388
c	146	141.2	14.1 239057	2	AC115637
	147	141.2	14.1 241726	2	AC097121
	148	141	14.1 231108	2	AC141480
c	149	141	14.1 240384	2	AC097041
c	150	140.8	14.1 220897	2	AL954635
	151	140.6	14.1 230183	2	AC107559
	152	140	14.0 222256	2	AC134482
	153	139.8	14.0 90406	2	AC133297
	154	139.8	14.0 245248	2	AC117869
c	155	139.2	13.9 161644	9	AL161645
c	156	139	13.9 212977	2	AC134108
c	157	139	13.9 256759	2	AC126002
	158	138.6	13.9 160553	5	AL807792
	159	138.6	13.9 201679	2	AC133459
	160	138.6	13.9 212253	2	AC140184
c	161	137.6	13.8 219108	2	AC095355
c	162	137.4	13.7 39723	9	AC141930
c	163	137.4	13.7 194204	2	AC131531
c	164	137.4	13.7 219486	2	AC116283
c	165	137	13.7 192613	2	AC141326
c	166	136.8	13.7 252686	2	AC106177
c	167	136.4	13.6 152589	2	BX510343
c	168	136.4	13.6 265229	2	AC121468
c	169	136.2	13.6 313488	2	AC128784
	170	135	13.5 181326	10	AL929165
	171	135	13.5 199495	2	AC079218
	172	135	13.5 238039	2	AC140796
	173	134	13.4 146824	2	AC118745
c	174	134	13.4 212899	2	AC102580
	175	133.8	13.4 224024	10	AL662802
c	176	133.8	13.4 230310	2	AC120079
	177	133.6	13.4 253951	2	AC128478
	178	133	13.3 228588	2	AC096241
c	179	132.8	13.3 265401	2	AC099177
c	180	132.6	13.3 199474	10	AC122873
c	181	131.8	13.2 164138	2	AC109432
c	182	131.8	13.2 178273	2	AC005308
	183	131.8	13.2 253132	3	AE014846
	184	131.6	13.2 231921	2	AC127906
	185	131.4	13.1 204317	2	AC119659
	186	131.4	13.1 293976	2	AC133832
c	187	131	13.1 163531	2	AL929518
	188	131	13.1 196947	10	AC084826
	189	131	13.1 261251	2	AC098618
c	190	130.8	13.1 181376	2	AC114990
c	191	130.8	13.1 238725	2	AC130214
	192	130.4	13.0 211758	2	AC095451
c	193	130.2	13.0 205843	2	AC141957
	194	129.4	12.9 250264	2	AC108970
	195	129	12.9 154918	9	AL138781
c	196	128.8	12.9 110000	2	AC113631_0
c	197	128.6	12.9 132948	2	AC006410
c	198	128.4	12.8 77436	2	AC004581
	199	128.4	12.8 147854	9	AC021422
c	200	128.4	12.8 179236	2	AC005143
c	201	128	12.8 224052	2	AC124316
	202	128	12.8 292785	2	AC134591
	203	128	12.8 321062	2	AC118593
c	204	127.6	12.8 35100	6	AR065849
c	205	127.6	12.8 35100	6	AR127847
c	206	127.6	12.8 35100	6	AR194749
	207	127.6	12.8 68595	2	AC084772
c	208	127.6	12.8 137508	14	KSU75698
c	209	127.4	12.7 201328	10	AL840635
	210	127.4	12.7 241394	2	AC131549
	211	127.4	12.7 261236	2	AC119381

AC115960 Mus muscu
AC119002 Rattus no
AC140324 Mus muscu
AC120675 Rattus no
AC132173 Rattus no
AC095908 Rattus no
AC138388 Mus muscu
AC115637 Rattus no
AC097121 Rattus no
AC141480 Mus muscu
AC097041 Rattus no
AL954635 Homo sapi
AC107559 Rattus no
AC134482 Rattus no
AC133297 Rattus no
AC117869 Rattus no
AL161645 Human DNA
AC134108 Rattus no
AC126002 Rattus no
AL807792 Zebrafish
AC133459 Mus muscu
AC140184 Mus muscu
AC095355 Rattus no
AC141930 Homo sapi
AC131531 Rattus no
AC116283 Rattus no
AC141326 Rattus no
AC106177 Rattus no
BX510343 Danio rer
AC121468 Rattus no
AC128784 Rattus no
AL929165 Mouse DNA
AC079218 Mus muscu
AC140796 Mus muscu
AC118745 Mus muscu
AC102580 Mus muscu
AL662802 Mouse DNA
AC120079 Rattus no
AC128478 Rattus no
AC096241 Rattus no
AC099177 Rattus no
AC122873 Mus muscu
AC109432 Rattus no
AC005308 Plasmodiu
AE014846 Plasmodiu
AC127906 Rattus no
AC119659 Rattus no
AC133832 Rattus no
AL929518 Danio rer
AC084826 Genomic s
AC098618 Rattus no
AC114990 Mus muscu
AC130214 Mus muscu
AC095451 Rattus no
AC141957 Rattus no
AC108970 Rattus no
AL138781 Human DNA
AC113631 Rattus no
AC006410 Homo sapi
AC004581 Homo sapi
AC021422 Homo sapi
AC005143 Homo sapi
AC124316 Mus muscu
AC134591 Mus muscu
AC118593 Mus muscu
AR065849 Sequence
AR127847 Sequence
AR194749 Sequence
AC084772 Homo sapi
U75698 Kaposi's sa
AL840635 Mouse DNA
AC131549 Rattus no
AC119381 Rattus no

c 212	127.2	12.7	284984	2	AC107121	AC107121 Rattus no
c 213	126.8	12.7	189805	9	AC004169	AC004169 Homo sapi
214	126.8	12.7	250286	2	AC119640	AC119640 Rattus no
c 215	126.6	12.7	190987	2	AC113883	AC113883 Rattus no
c 216	126.6	12.7	254372	2	AC132548	AC132548 Rattus no
c 217	126.2	12.6	111023	9	AL592464	AL592464 Human DNA
c 218	126.2	12.6	184104	9	AC020709	AC020709 Homo sapi
219	126.2	12.6	266712	2	BX088533	BX088533 Danio rer
220	126	12.6	168153	2	AC103388	AC103388 Mus muscu
221	125.8	12.6	565	11	BV015874	BV015874 S208P6661
222	125.4	12.5	159117	2	AC132272	AC132272 Mus muscu
c 223	125.4	12.5	161984	9	AC106772	AC106772 Homo sapi
224	125.2	12.5	41951	2	AC139177	AC139177 Ciona sav
c 225	125	12.5	120084	9	AC006205	AC006205 Homo sapi
c 226	125	12.5	201702	10	AL772335	AL772335 Mouse DNA
227	125	12.5	231877	2	AC121634	AC121634 Rattus no
c 228	125	12.5	239099	2	AC108535	AC108535 Rattus no
c 229	124.4	12.4	276624	2	AC132560	AC132560 Rattus no
c 230	123.8	12.4	113891	5	AL713987	AL713987 Zebrafish
231	123.8	12.4	240525	2	AC096895	AC096895 Rattus no
232	123.6	12.4	220015	2	AC123418	AC123418 Rattus no
c 233	123.6	12.4	231698	2	AC118117	AC118117 Rattus no
234	123.6	12.4	247380	2	AC127159	AC127159 Rattus no
235	123.6	12.4	260202	2	AC114366	AC114366 Rattus no
c 236	123.4	12.3	84563	5	AC096885	AC096885 Danio rer
c 237	123.4	12.3	184986	2	BX248111	BX248111 Danio rer
c 238	123.4	12.3	193067	2	BX323458	BX323458 Danio rer
c 239	123.2	12.3	179192	2	AC044848	AC044848 Mus muscu
240	123.2	12.3	220050	10	AL583887	AL583887 Mouse DNA
c 241	123.2	12.3	236344	10	AL583889	AL583889 Mouse DNA
242	123.2	12.3	236394	2	AC120124	AC120124 Mus muscu
c 243	123	12.3	112085	9	AC022132	AC022132 Homo sapi
c 244	123	12.3	112085	9	AC026712	AC026712 Homo sapi
c 245	123	12.3	139111	9	AC008768	AC008768 Homo sapi
c 246	123	12.3	183330	2	AC008159	AC008159 Homo sapi
c 247	123	12.3	239920	2	AC123504	AC123504 Rattus no
248	122.8	12.3	238653	2	AC099453	AC099453 Rattus no
c 249	122.8	12.3	251506	2	AC106244	AC106244 Rattus no
250	122.8	12.3	263700	2	AC098196	AC098196 Rattus no
251	122.6	12.3	222259	2	AC110858	AC110858 Rattus no
252	122.6	12.3	231864	2	AC097246	AC097246 Rattus no
253	122.6	12.3	261603	2	AC098318	AC098318 Rattus no
c 254	122.4	12.2	198147	2	AC115032	AC115032 Mus muscu
255	122.2	12.2	233962	2	AC094754	AC094754 Rattus no
c 256	122	12.2	86165	5	AL954681	AL954681 Zebrafish
c 257	121.6	12.2	131112	9	AL157712	AL157712 Human DNA
c 258	121.6	12.2	174231	9	AC021054	AC021054 Homo sapi
c 259	121.6	12.2	185500	2	AC118940	AC118940 Mus muscu
260	121.6	12.2	208049	2	AC108342	AC108342 Rattus no
261	121.2	12.1	110000	2	AC110828_3	Continuation (4 of
c 262	121.2	12.1	175581	9	AP001525	AP001525 Homo sapi
c 263	121.2	12.1	185603	9	AP001086	AP001086 Homo sapi
264	121.2	12.1	286029	2	AC110845	AC110845 Rattus no
265	121.2	12.1	310991	2	AC109011	AC109011 Rattus no
c 266	121	12.1	179306	2	AC130070	AC130070 Rattus no
267	121	12.1	180485	10	AL683823	AL683823 Mouse DNA
268	120.8	12.1	219092	2	AC136030	AC136030 Rattus no
269	120.8	12.1	258815	2	AC127817	AC127817 Rattus no
270	120.2	12.0	188994	2	AC110325	AC110325 Rattus no
271	120.2	12.0	305997	2	AC123003	AC123003 Rattus no
272	120	12.0	180846	2	AC145151	AC145151 Homo sapi
273	120	12.0	212419	2	AC099586	AC099586 Mus muscu
274	119.6	12.0	175636	9	AC021744	AC021744 Homo sapi
275	119.6	12.0	197014	2	AC124133	AC124133 Mus muscu
c 276	119.6	12.0	240279	2	AC126486	AC126486 Rattus no
c 277	119.6	12.0	270456	2	AC111575	AC111575 Rattus no
c 278	119.4	11.9	172485	2	AC141978	AC141978 Rattus no
c 279	119.4	11.9	244521	2	AC125996	AC125996 Rattus no
280	119.4	11.9	246075	2	AC111469	AC111469 Rattus no
281	119	11.9	172533	2	AC120551	AC120551 Mus muscu
c 282	118.6	11.9	87233	2	AC141102	AC141102 Rattus no
283	118.6	11.9	124636	9	AP000721	AP000721 Homo sapi
284	118.6	11.9	252349	2	AC097200	AC097200 Rattus no
c 285	118.6	11.9	271726	2	AC129344	AC129344 Rattus no
286	118.4	11.8	239392	2	AC115417	AC115417 Rattus no
287	118.2	11.8	107103	9	AL589702	AL589702 Human DNA
c 288	118	11.8	168949	2	AC139895	AC139895 Rattus no
c 289	118	11.8	181194	2	AC108444	AC108444 Mus muscu
290	118	11.8	190670	2	AC122103	AC122103 Rattus no
291	118	11.8	221568	2	AC103452	AC103452 Rattus no
292	118	11.8	233663	2	AC115486	AC115486 Rattus no
c 293	118	11.8	252813	2	AC118528	AC118528 Rattus no
294	118	11.8	253978	2	AC094368	AC094368 Rattus no
295	117.8	11.8	1360	6	AX646913	AX646913 Sequence
c 296	117.8	11.8	182945	2	AC069004	AC069004 Homo sapi
c 297	117.8	11.8	258510	2	AC130514	AC130514 Rattus no
298	117.4	11.7	173641	2	AC141089	AC141089 Rattus no
299	117.2	11.7	183111	2	AC099708	AC099708 Mus muscu
300	117	11.7	150612	2	AC126099	AC126099 Rattus no

ALIGNMENTS

RESULT 1
AF083501
LOCUS AF083501 133719 bp DNA linear VRL 26-JUN-2000
DEFINITION Macaca mulatta rhadinovirus 17577, complete genome.
ACCESSION AF083501
VERSION AF083501.3 GI:8714565
KEYWORDS .
SOURCE Macaca mulatta rhadinovirus 17577
ORGANISM Macaca mulatta rhadinovirus 17577
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Rhadinovirus.
REFERENCE 1 (bases 1 to 133719)
AUTHORS Searles,R.P., Bergquam,E.P., Axthelm,M.K. and Wong,S.W.
TITLE Sequence and genomic analysis of a Rhesus macaque rhadinovirus with similarity to Kaposi's sarcoma-associated herpesvirus/human herpesvirus 8
JOURNAL J. Virol. 73 (4), 3040-3053 (1999)
MEDLINE 99174001
PUBMED 10074154
REFERENCE 2 (bases 1 to 133719)
AUTHORS Searles,R.P., Bergquam,E.P., Axthelm,M.K. and Wong,S.W.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-1998) Division of Pathobiology and Immunology, Oregon Health Sciences University/ Oregon Regional Primate Research Center, 505 NW I85th Avenue, Beaverton, OR 97006, USA
REFERENCE 3 (bases 1 to 133719)
AUTHORS Searles,R.P., Bergquam,E.P., Axthelm,M.K. and Wong,S.W.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-1999) Division of Pathobiology and Immunology, Oregon Health Sciences University/ Oregon Regional Primate Research Center, 505 NW I85th Avenue, Beaverton, OR 97006, USA
REMARK Sequence update by submitter
REFERENCE 4 (bases 1 to 133719)
AUTHORS Searles,R.P., Bergquam,E.P., Axthelm,M.K. and Wong,S.W.
TITLE Direct Submission
JOURNAL Submitted (26-JUN-2000) Division of Pathobiology and Immunology, Oregon Health Sciences University/ Oregon Regional Primate Research Center, 505 NW I85th Avenue, Beaverton, OR 97006, USA
REMARK Sequence update by submitter
COMMENT On Jun 26, 2000 this sequence version replaced gi:5508840.
FEATURES Location/Qualifiers
source 1. .133719
/organism="Macaca mulatta rhadinovirus 17577"
/mol_type="genomic DNA"
/db_xref="taxon:83534"
repeat_region 1. .840
/note="partial terminal repeat"
CDS 1353. .2624
/codon_start=1
/product="R1"
/protein_id="AAD21330.1"
/db_xref="GI:4494908"

/translation="MFVLVLFMLLQPVSVELLPAKLTSVPTWCPPHPGDTYLLTCRGT
STARDQRSTQWFRNNTLMRGSNFYGRLVSVTPNATISDRYACQTKTTTRSNNIDFRVS
SSRLTLQERCSSYGYTYANNTRVLRCYSGGNVTLRNVVFHLNGTAVINGTTTNIHTFV
LTEKTGGTYFCSAFIGNEKFYSQTINVFFTSFTFKPTNDIPNESHFNKTGQIQQTASV
QHPENYVVFSVPVFSIGVLTGIAISLIMCWLFTIRCNENSESSTNSYASQTSYIQPSH
NQRSNTNECSRHTYRNAHQEESIEELPNQHTSETDSCCQLVLLEVKNVAYDGPQENTI
NEVMEQYDDVVVKNIEQTSYEDNVEHMDYSDTINPNFNYYSGLILEEVDEVFYNELEN
QYHGLILENLDHNEYNHLNELNMIEQYDWLE"

CDS complement(2692. .3258)
/note="ORF 2; similar to Kaposi's sarcoma-associated herpesvirus ORF 2"
/codon_start=1
/product="dihydrofolate reductase"
/protein_id="AAD21331.1"
/db_xref="GI:4494909"
/translation="MDITVNCIVAVDEQLGIGKNGTMPWPYLRNEMMYFQKMTSTPSV
VGEKNVVIMGKRTWFSIPEKKRPLVNRINIILSRELREPPHGAHFLARTLDDAFNFYR
QYKLKEQLNTVWVIGGKSVYESVLNYKCPLKLYITRIMESFDCDVFFPSINFTEYTML
SEIPGKDTNFEENGIKYKFQVYEKNFNK"

CDS 3676. .5613
/note="ORF 4; similar to Kaposi's sarcoma-associated herpesvirus ORF 4"
/codon_start=1
/product="complement binding protein"
/protein_id="AAD21332.1"
/db_xref="GI:4494910"
/translation="MTFKLFPLFLLHAIMYVHCDENCKPPHFTEYRVKSNTEKDLYSV
GETAELICRPGYVTNTKIITTECLQNGTWSTPNFPCDRKRCPTPADLLNGAVHIHGGD
NALKFGSNISYECNEGYDLIGSNVRFCILQDTENVNWDSNEPVCEIQKCIKPPAVEHG
DYLPNQDVYNYGDAITFKCSLSYTLVGSTTLVCTSNKKWSNSFPTCLMLVCESPQIDN
GYIDIGLSRRYNHGQSITVKCSDGYNIVGPETLTCTNTTWVPPLPKCVLVTNNPSTPM
PETPMPETPTPDYQKINLSTAKTATTPNAFVTTVVSPEKDDVTCVKPHFERFMVKAEN
DKEKYSVGASVELICRPGFTKMQSTVSVECLSNGTWTAPNAKCHRKKCPTPQELLNGE
YIVTSGEDAFKYGTNITYKCNEGYQLLGSMVRICMLKDDLKTVDWEPKAPICDIEKCK
PPPQITNGKYHPVKDFYQYLDTVTFSCNRDFSLVGDEMTTCISNTWNKPFPRCEQITC
SAPNIAHGKLLTGSSSVYKYGQSVTIGCETGFTLIGSEISTCKDSSWDPPLPTCVPAV
SMPSDTPKPETKKPNTPTPEAPKPNTPNVGTHTPFKPPPQNPPIAPPMSKWKRHVVLV
LFASVASLLFVLAALYCCFLK"

CDS 6045. .9443
/note="ORF 6; similar to Kaposi's sarcoma-associated herpesvirus ORF 6"
/codon_start=1
/product="ssDNA binding protein"
/protein_id="AAD21333.1"
/db_xref="GI:4494911"
/translation="MASKGNAGQPLEDNQGSRAPIGACGYVYAYSKQDFPFAEASILG
NRPSGSGVFSLPILYGLTVEHEFPLTVKAAYKKVDTTTLAVKVTCFHREVIVFHNASL
FRPVFDGTGLNELCEEARALFGYTQFIEPGPPHSIWNPLECPQLPDKDEMFLGVVVTE
GFKERLWRGCLVPAVFQTQQVQIAGRQAFKVPLYDEDLFAPHGHRMPRFYHKDVSAYL
YDSLFTSIAQALRLKDVTAVIHATEKQFMQDHYKIAKIVQAKQFSTTLPKTTDGSSHM
IVDSVVAELALSYGCMFLECPQDACELLNYDSWPIFDGCDSPEARVNALERWSAEQAV
HVAGQLFAANSVLYLTKVQKQAPRGQKGDVNVYNSFFLQHGLGFLNEATIKENGSEAF
KGVPSNALDGSSFTPYHLAYAASFSPHLLAKLCYYMQFLQHHKSSTNQAFNMVHYVGT
AANSEMCTLCHGNTPATCLNTLFYRLKDRFPAVTTPQRRDPYVVTGTAGTFNDLEILG
NFASFRDREEDGNPADEHPKYTYWQLCQTVTEKLSAIGITEDHDNHVNLITNIQSFLR
VFKGIDSIVDGEVMKFVNSMIKNNFNFREHVKSVHHILQFCCNVYWQAPCAVFLNLYY
KSLLWIIQDICLPYCMIYEQDNPAMGILPSEWLKMHFQTLWTNFKAACLDRGVLTGCE
LKIVHRDMFCDFFDTDAGSNGLMAPFKMQVRIARAMMVVPKSIKIKNRIIFSNTAGSE
AVQSGFVKPTGTRDTYVVAGPYMKFLNSLHRALFPDTKTAALYLWHKISQTNKTPVLK
DVPDDELAELVSYVKTNSLAFEETNVLDVVPDSLMSYARIKLNGAILRACGQIQFYAT
TLHCLTPVLQTIDAEEYPHVLGSAAIATPVAYLAEIRGRTALTVQTTARQPVAATGRL
RPVITVPMVVNKYTGVNGNNNVFHCGNLGYFAGRGVDRNLWPESSPFKKTGVSAMLRK
RHVMMTPIIDRLIKRAAGQTISTFEAESVKRSVQALLEDKDNPNLLKSVILELIRHLG
KGCQDLSSEDVQYYLGDYCMLTDEVLFTLDNIAQSGVPWTIEDAGALIEDRQDADDLQ
FVDSDDIATASCQPPEEQLPTPSAGALLAGKKRKINALLSDLDL"

CDS 9468. .11528
/note="ORF 7; similar to Kaposi's sarcoma-associated herpesvirus ORF 7"
/codon_start=1
/product="transport protein"
/protein_id="AAD21334.1"
/db_xref="GI:4494912"
/translation="MARELAALYAQLSALAVDLSLVIFADPRSIDGARILKTKTQIEN
LNRDLLPLLREQNSVETSSLSLEVEHLAKNIEDKLGELERSLRQRYSSREHFETLHLR
PECHYHSTVTFQFYGGGLIDVNMCLINDVELLCKRLGSVFYCIGANEALSGLNRVLTF
LSTLRGISPIPHPDLYVTSVPCVQCLREIELVPNQGSSLLAVLADRHCDHLCKKVRAE
PIHGLFETELSQLGLKVTKRSDATQHGVRSSADQLRESSLAAIQDHNIFKRVSASIME
LSNLIYWNAGQTGLQTGTENECSQMARLLTHEADMHEHRALITPKLSATHFYDCFRPD
PIESLFCGGLFNSIDDTINALSRDCSVTFFQQANYTNVMRKQNELFTRLNSILRQGSA
GSQKPATPSEPRTTTVAATAASDVIKDAQYRKEQYMKKVARDGFKKLTECLQTQSAVL
ANALCMRVWGGVAYGEASELVNHFLLRRRFVALPWEARCRSDQILFENSKYIKNSLYS
QRLSREHVEIITLQFYGLITGPLTRQSDLFPGPANVALAQCFEAAGMLPHHKMLVSEM
IWPQIQPKDWIDQTFNRFYQLPEGDLNAVQKSAWCFIRELVLSVALYNRTWEKTLRIF
SLAREKLSISNLDVKGLTSGLYLTYEQDAPLVLISQNTGWIFKDLYALLYHHLQLSDG
HDDN"

CDS 11515. .14004
/note="ORF 8; similar to Kaposi's sarcoma-associated herpesvirus ORF 8"
/codon_start=1
/product="glycoprotein B"
/protein_id="AAD21335.1"
/db_xref="GI:4494913"
/translation="MMITNRTRRLLRAWVVIIAIGTAVGENVTTPKGATTTAKPTPGP
STPTPPENPPRAEAFKFRVCSASATGELFRFNLEKTCPGTEDKTHQEGILMVFKKNIV
PHIFKVRRYRKVATSVTVYRGWTETAVTGKQEVIRPVPQYEINHMDTTYQCFSSMRVN
VNGIVNTYTDRDFTNQTVFLQPVEGLTDNIQRYFSQPVLYTTPGWFPGIYRVRTTVNC
EIVDMIARSAEPYSYFVTALGDTVEVSPFCHNDSTCSVAEKTENGLGARVLTNYTIVD
FATRQPTTETRVFADSGEYTVSWKAEDPKSAVCALTLWKTFPRAIQTTHEASYHFVAN
DVTATFTSPLSQVTNFTGTYPCLNDVIQKTLNATIKKLSDTHATNGSEQYYETEGGLF
LLWQPLTPLSLADEMRELNGTTPAPPTTTSTANRVRRSVGTNEQATDDLAAPQLQFAY
DKLRASINKVLEELSRAWCREQVRDTYMWYELSKINPTSVMTAIYGRPVSAKFVGDAI
SVTDCVAVDQASVSIHKSLRTSTPGICYSRPPVTFRFLNSTTLFKGQLGPRNEIILTD
NQVEACKETCEHYFIASNVTYYYKDYVFVKKINTSEISTLGTFIALNLSFIENIDFRV
IELYSRAEKKLSGSVFDIETMFREYNYYTQRLAGLREDLDNTIDLNRDRLARDLSEIV
ADLGDVGRTVVNVASSVITLFGSIVSGFINFIKSPFGGMLMILVIVAVVLIVFALNRR
TNAIAQAPIRMIYPDIDKMQPSGGKVDQEQIKNILAGMHQLQQEERRRLDEQQRSAPS
LFRRASDGLKRRFRGYKPLENEEAQEYEMSK"

CDS 14122. .17166
/note="ORF 9; similar to Kaposi's sarcoma-associated herpesvirus ORF 9"
/codon_start=1
/product="DNA polymerase"
/protein_id="AAD21336.1"
/db_xref="GI:4494914"
/translation="MDFFNPYLGPRGPRPHSHRGTDAPAPAGAGAVQPPPDVCRLIPA
CLRTPGAGGMIPVTIPFPPPTYFENGARGDVLLANERSMWTARDRKPVAPDPQDQSITF
HAYDVVETTYAADRCAEVPSRFQTDIIPSGTVLKLLGRTEDGTSVCVNVFRQQVYFYA
KVPAGINVTHILQQALKNTAGRAACGFSTRRVNKRILKTYDVAEHPVTEITLSSGSML
STLSDRLVACGCEVFESNVDAVRRFVLDHGFTTFGWYSCARATPRLAARDARTALEFD
CSWEDLSVQADRSDWPPYRIVAFDIECTGEAGFPCATRDGDAVIQISCVFYTTREGAP
NPPNILFSVGTCDPIPDTDVLEFPSEYDMLVSFFAMIRDFEVDFLTGYNISNFDLPYL
ITRASQVYNLRLNEYTKIKTGSIFEVHEPRGGGGGFMRSVSKIKIAGIVPIDMYQVCR
EKLSLSDYKLDTVARQCLGGKKEDVSYKDIPPLFRSGPGGRAKVGSYCVMDSVLVMDL
LKMFMIHVEISEIAKLAKIQARRVLTDGQQLRVFSCLLEAAARENFILPVPTPEGQGG
YQGATVINPIPGFYDEPVLVVDFASLYPSIIQAHNLCYSTMIHGRDLHLHPNLTPDDY
ETFVLSGGPVHFVKKHKRESLLGRLLTVWLEKRRAIRRTLAACDDPSLKTILDKQQLA
IKVTCNAVYGFTGVASGLLPCINIAETVTLRGRTMLEMSKSYVEALTTEDLRTRLGRE
VTARHGARFRVVYGDTDSLFIACDGYSAEAVSAFCDDLAARITADLFPPPIKLEAEKT
FKCLLLLTKKRYIGVLLNDKMVMKGVDLIRKTACKFVQERCRAILDLVLHDPEVKAAA
RLLCKRPPHAVYEEGLPAGFIKIVEVLNASYLDLRNSVVPIEQLTFSTELSRPVCDYK
TTNLPHLAVYQKLASRCEELPQVHDRIPYVFVDAPGSLKSDLAEHPDYVRQHQIPVAV
DLYFDKLVHGAANILQCLFGNNADTTVAILYNFLNVPYKLFS"

CDS 17261. .18511
/note="ORF 10; similar to Kaposi's sarcoma-associated herpesvirus ORF 10"
/codon_start=1
/product="unknown"
/protein_id="AAD21337.1"
/db_xref="GI:4494915"
/translation="MLVNELSVVLGDWEVTFHRGRFSFVNLTRLQTFKGHGGYARVRL
PFSLDQLLHQHFAFGLVTRLKELPPFSDCVALIAPLDSGGDADAARVAPGFVLDSSRP
LTVWVNASGRHTIRFCLLFLKPIDLERAVTYVFGENGGARSEGTPKPTCATESLPGGP
LRVSGEASQTSPHSFVAYFPTANSVACLSLLRLQVRPFSDDAAHRDARISPKYVTFSN
SGGNVCKASVHTLSPSRCKTAQMEIIYAPGDPNAEIVLGQSGPVLPTHTGGRVLGVYA
DAEKTIQPGSSAEVRVQLIFQQGAAARGDLAFLVTGVAPEPLFVVTPALLLSGCTTHL
RLFNPNGTPTTIKRDTLVAAAAPCPVVRLSSADDAPRDLVASPDTGALSINAFTIPVG

```
                FPGVVSAECHVSLRDNGVHERMNH"
  CDS           18520. .19749
                /note="ORF 11; similar to Kaposi's sarcoma-associated
                herpesvirus ORF 11"

  Query Match           100.0%;  Score 1000;  DB 14;  Length 133719;
  Best Local Similarity  100.0%;  Pred. No. 2.4e-224;
  Matches 1000;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACA 60

Qy         61 CCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 120

Qy        121 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCT 180

Qy        181 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCT 240

Qy        241 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 300

Qy        301 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 360

Qy        361 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 420

Qy        421 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 480

Qy        481 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 540

Qy        541 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 600

Qy        601 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 660

Qy        661 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGG 720

Qy        721 GGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGGGCTGCCGGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGGGCTGCCGGC 780

Qy        781 CAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACCCGGGCCTCTCCCCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACCCGGGCCTCTCCCCC 840

Qy        841 AGGCACAAAGCAGTAGGGCTGGCCAGGGGAAACAGTAAGGCTGATTGCTTGCTGAAAAAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AGGCACAAAGCAGTAGGGCTGGCCAGGGGAAACAGTAAGGCTGATTGCTTGCTGAAAAAC 900

Qy        901 AGTAGGGCTGCTGGTTTGTTGCTAACGGAAAAAGGGGAGGTGTGTGTATGCTTTTGTGAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTAGGGCTGCTGGTTTGTTGCTAACGGAAAAAGGGGAGGTGTGTGTATGCTTTTGTGAC 960
```

```
Qy        961 TTTCCAGCGGAGGGGGTACAGTGCACGGCCAAGTTACAAG 1000
              ||||||||||||||||||||||||||||||||||||||||
Db        961 TTTCCAGCGGAGGGGGTACAGTGCACGGCCAAGTTACAAG 1000


RESULT 2
AC116487/c
LOCUS       AC116487              148106 bp    DNA     linear   HTG 12-JUL-2002
DEFINITION  Mus musculus clone RP23-428I9, WORKING DRAFT SEQUENCE,. 22 ordered
            pieces.
ACCESSION   AC116487
VERSION     AC116487.2  GI:21735111
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 148106)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-428I9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 148106)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 148106)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 12, 2002 this sequence version replaced gi:19774593.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
```

```
            ---------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ---------- Project Information
                Center project name: L23947
                Center clone name: 428_I_9
            ---------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 138537 bases at least Q40
                Consensus quality: 143163 bases at least Q30
                Consensus quality: 144845 bases at least Q20
                Insert size: 160000; agarose-fp
                Insert size: 146006; sum-of-contigs
                Quality coverage: 5.6 in Q20 bases; agarose-fp
                Quality coverage: 6.1 in Q20 bases; sum-of-contigs
            ----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 22 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    13805: contig of 13805 bp in length
            *    13806    13905: gap of 100 bp
            *    13906    14560: contig of 655 bp in length
            *    14561    14660: gap of 100 bp
            *    14661    15529: contig of 869 bp in length
            *    15530    15629: gap of 100 bp
            *    15630    16744: contig of 1115 bp in length
            *    16745    16844: gap of 100 bp
            *    16845    18432: contig of 1588 bp in length
            *    18433    18532: gap of 100 bp
            *    18533    19971: contig of 1439 bp in length
            *    19972    20071: gap of 100 bp
            *    20072    21780: contig of 1709 bp in length
            *    21781    21880: gap of 100 bp
            *    21881    23204: contig of 1324 bp in length
            *    23205    23304: gap of 100 bp
            *    23305    24685: contig of 1381 bp in length
            *    24686    24785: gap of 100 bp
            *    24786    28365: contig of 3580 bp in length
            *    28366    28465: gap of 100 bp
            *    28466    30918: contig of 2453 bp in length
            *    30919    31018: gap of 100 bp
            *    31019    34321: contig of 3303 bp in length
            *    34322    34421: gap of 100 bp
            *    34422    36281: contig of 1860 bp in length
            *    36282    36381: gap of 100 bp
            *    36382    39099: contig of 2718 bp in length
            *    39100    39199: gap of 100 bp
            *    39200    42684: contig of 3485 bp in length
            *    42685    42784: gap of 100 bp
            *    42785    46395: contig of 3611 bp in length
            *    46396    46495: gap of 100 bp
            *    46496    51682: contig of 5187 bp in length
            *    51683    51782: gap of 100 bp
            *    51783    56221: contig of 4439 bp in length
            *    56222    56321: gap of 100 bp
            *    56322    66418: contig of 10097 bp in length
            *    66419    66518: gap of 100 bp
            *    66519    83479: contig of 16961 bp in length
            *    83480    83579: gap of 100 bp
            *    83580   110083: contig of 26504 bp in length
            *   110084   110183: gap of 100 bp
            *   110184   148106: contig of 37923 bp in length.
FEATURES             Location/Qualifiers
     .source         1. .148106
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-428I9"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .13805
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    13906. .14560
                     /note="assembly_fragment"
     misc_feature    14661. .15529
                     /note="assembly_fragment"
     misc_feature    15630. .16744
                     /note="assembly_fragment"
     misc_feature    16845. .18432
                     /note="assembly_fragment"
     misc_feature    18533. .19971
                     /note="assembly_fragment"
     misc_feature    20072. .21780
                     /note="assembly_fragment"
     misc_feature    21881. .23204
                     /note="assembly_fragment"
     misc_feature    23305. .24685
                     /note="assembly_fragment"
     misc_feature    24786. .28365
                     /note="assembly_fragment"
     misc_feature    28466. .30918
                     /note="assembly_fragment"
     misc_feature    31019. .34321
                     /note="assembly_fragment"
     misc_feature    34422. .36281
                     /note="assembly_fragment"
     misc_feature    36382. .39099
                     /note="assembly_fragment"
     misc_feature    39200. .42684
                     /note="assembly_fragment"
     misc_feature    42785. .46395
                     /note="assembly_fragment"
     misc_feature    46496. .51682
                     /note="assembly_fragment"
     misc_feature    51783. .56221
                     /note="assembly_fragment"
     misc_feature    56322. .66418
                     /note="assembly_fragment"
     misc_feature    66519. .83479
                     /note="assembly_fragment"
     misc_feature    83580. .110083
                     /note="assembly_fragment"
     misc_feature    110184. .148106
                     /note="assembly_fragment"
BASE COUNT    36315 a  37882 c  38018 g  33788 t   2103 others
ORIGIN

  Query Match             28.7%;  Score 286.6;  DB 2;  Length 148106;
  Best Local Similarity   64.9%;  Pred. No. 4.6e-57;
  Matches  424;  Conservative    0;  Mismatches  229;  Indels    0;  Gaps    0;

Qy         65 GTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 124
              ||  | ||| |||  || | | | |  || |||||     |   |   || ||||||||
Db      40780 GTCTGTCTGTCTGTTTGTTTGTTTGTTTGTTTGCTTGCTTGCTTGCCTGCTTGCTTGCTT 40721

Qy        125 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 184
              ||  || ||||||||     |   |   || ||| ||||     |   |   || |||||
Db      40720 GCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTT 40661

Qy        185 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTC 244
              ||| ||  || ||||||||     | | |   |  |||| ||||||| |   | |||||
Db      40660 GCTTGCCTGCTTGCTTGCTTGCCTGCCTGCCTGCTTGCTTGCTTGCTTGCTTGCCTGCTT 40601

Qy        245 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 304
              ||| |||||||||| |||| ||| || ||||||||||  ||  || ||||||| ||||
```

```
Db      40600 GCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCCTGCCTGCTTGCTTGCCTGCTT 40541

Qy        305 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 364
              ||| |||||| |||||||| ||| || ||| |||||||| ||| ||||||||||||||||
Db      40540 GCTTGCCTGCCTGCTTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTT 40481

Qy        365 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 424
              ||| || |||||||||||| ||| ||||||||||||||| ||  || ||||||||
Db      40480 GCTTGCTTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCCT 40421

Qy        425 GACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAG 484
              |   |   | ||||||||||||||     |   |   ||||||||| | ||||| | |
Db      40420 GCTTGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCCTGCCTGCTTGCTTGCTTGCTT 40361

Qy        485 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 544
              |  |||| |  |||| |   |  |||| |  |||| |   | ||||| |  |||  |
Db      40360 GCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGCCTGCTTGCTTGCCTGCTT 40301

Qy        545 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 604
              |  |||| | ||||| |   | || |  |  |||| |   |  |||| |  |||| |
Db      40300 GCTTGCTTGCCTGCTTGCTTGCCTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 40241

Qy        605 GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 664
              | || || |  |||| |   | ||||| |  |||| |   |  |||| |  |||| |
Db      40240 GCCTGCTTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCTTGCTT 40181

Qy        665 GGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAG 717
              |  |||| | ||||| |   |  |||||||| |||||||||||||| |||| |
Db      40180 GCTTGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCTTGCCTGCTTG 40128
```

RESULT 3
AC099177
LOCUS AC099177 265401 bp DNA linear HTG 10-MAY-2003
DEFINITION Rattus norvegicus clone CH230-112E16, *** SEQUENCING IN PROGRESS ***, 2 unordered pieces.
ACCESSION AC099177
VERSION AC099177.6 GI:30522052
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 265401)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 265401)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 265401)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 10, 2003 this sequence version replaced gi:25139860.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
-------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
Center project name: GHAA
Center clone name: CH230-112E16
-------------- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 219762 bases at least Q40
Consensus quality: 223776 bases at least Q30
Consensus quality: 226617 bases at least Q20
Estimated insert size: 230040; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```
          * be preserved.
          *       1   263276: contig of 263276 bp in length
          *  263277   263376: gap of unknown length
          *  263377   265401: contig of 2025 bp in length.
FEATURES             Location/Qualifiers
     source          1. .265401
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-112E16"
     misc_feature    1. .1265
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    12027. .12919
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH365978"
     misc_feature    90541. .93662
                     /note="wgs_contig"
     misc_feature    136482. .137827
                     /note="wgs_contig"
     misc_feature    261732. .263276
                     /note="wgs_contig"
BASE COUNT    64515 a  48596 c  48625 g  66669 t  36996 others
ORIGIN

  Query Match             28.1%;  Score 280.6;  DB 2;  Length 265401;
  Best Local Similarity   62.4%;  Pred. No. 1.1e-55;
  Matches  439;  Conservative    0;  Mismatches  264;  Indels    0;  Gaps    0;

Qy         44 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 103
Db     138985 TGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCT 139044

Qy        104 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 163
Db     139045 CGCTTGCCTGCCTGCTTGCTTGCCTGCCTGCCTGCTTGCCTGCCTGCCTGCTTGCCTGCC 139104

Qy        164 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 223
Db     139105 TGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCTCGCTTGCCTGCC 139164

Qy        224 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 283
Db     139165 TGCTTGCTTGCCTGCCTGCCTGCTTGCCTGCTTGCTCACTTGCCTGCTTGCTTGCCTGCC 139224

Qy        284 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 343
Db     139225 TGCCTGCCTGCTTGCCTGCCTGCTTGCTTGTTCACTTGCCTGCTTGCCTGCTTGCTTGCT 139284

Qy        344 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403
Db     139285 TGCTCGCTTGCCTGCCTGCCTGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCT 139344

Qy        404 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCC 463
Db     139345 CGCTTGCCTGCCTGCCTGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCC 139404

Qy        464 TGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 523
Db     139405 TGCCTGCTTGCTCGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCT 139464

Qy        524 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTA 583
Db     139465 TGCCTGCTTGCTCGCTTGCCTGCTTGCTTGCTTGCCTGCCTGCCTGCTTGCCTGCTTGCT 139524

Qy        584 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 643
Db     139525 CACTTGCCTGCTTGCTTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCTTGCTTGTTCGCT 139584

Qy        644 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCT 703
Db     139585 TGCCTGCTTGCTTGCTTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCCTGCCTGCCTGCT 139644

Qy        704 TGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGC 746
Db     139645 TGCCTGCTTGCTCGCTTGCCTGCCTGCCTGCTTGCCTGCCTGC 139687


RESULT 4
AC119239
LOCUS       AC119239              165180 bp    DNA     linear   HTG 10-MAR-2003
DEFINITION  Mus musculus clone RP24-228F21, WORKING DRAFT SEQUENCE, 6 unordered
            pieces.
ACCESSION   AC119239
VERSION     AC119239.3  GI:28894656
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 165180)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-228F21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165180)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 165180)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
```

```
COMMENT     *On Mar 10, 2003 this sequence version replaced gi:28269570.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L25264
                Center clone name: 228_F_21
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 164030 bases at least Q40
                Consensus quality: 164471 bases at least Q30
                Consensus quality: 164601 bases at least Q20
                Insert size: 166000; agarose-fp
                Insert size: 164680; sum-of-contigs
                Quality coverage: 11.1 in Q20 bases; agarose-fp
                Quality coverage: 11.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1027: contig of 1027 bp in length
            *    1028     1127: gap of 100 bp
            *    1128    16448: contig of 15321 bp in length
            *   16449    16548: gap of 100 bp
            *   16549    76346: contig of 59798 bp in length
            *   76347    76446: gap of 100 bp
            *   76447   103034: contig of 26588 bp in length
            *  103035   103134: gap of 100 bp
            *  103135   153662: contig of 50528 bp in length
            *  153663   153762: gap of 100 bp
            *  153763   165180: contig of 11418 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165180
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-228F21"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .1027
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    1128. .16448
                     /note="assembly_fragment"
     misc_feature    16549. .76346
                     /note="assembly_fragment"
     misc_feature    76447. .103034
                     /note="assembly_fragment"
     misc_feature    103135. .153662
                     /note="assembly_fragment"
     misc_feature    153763. .165180
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    44313 a  40435 c  37850 g  42082 t    500 others
ORIGIN

  Query Match           27.3%;  Score 272.6;  DB 2;  Length 165180;
  Best Local Similarity  60.7%;  Pred. No. 8.8e-54;
  Matches  446;  Conservative    0;  Mismatches  289;  Indels    0;  Gaps    0;

Qy          44 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 103
               ||| ||||||    |     |   |   |  ||| ||| || |  || || |||||| |  |
Db      114820 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGGTTACTTGCTTGCTTGGTTGGT 114879

Qy         104 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 163
                 |   || |||||||| ||| || ||||||||     |   |   || ||||||||
Db      114880 TACTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 114939

Qy         164 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 223
                   |       || |||||||| ||| || ||||||||     |   |   |  ||||
Db      114940 TGCTTGGTTACTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 114999

Qy         224 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 283
               ||||||| |   |  |||| ||| || || || |||||| | |  | ||||||||||||
Db      115000 TGCTTGCTTGCTTGCTTGCTTGCTTGGTTACTTGCTTGGTTACTTGCTTGCTTGCT 115059

Qy         284 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 343
               ||| || |||||||||||| ||| || |||||||||||| ||| || || || ||||||
Db      115060 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGGTTACTTGCT 115119

Qy         344 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403
               ||| || |||||||||||| ||| || || || |||||| ||| || ||||||||||||
Db      115120 TGCTTGCTTGCTTGCTTGCTTGCTTGGTTACTTGCTTGCTTGCTTGCTTGCTTGCT 115179

Qy         404 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCC 463
               ||| || ||||||||     |   |  || | ||||||||||     |   |   ||
Db      115180 TGCTTGCTTGCTTGCTTGCTTGCTTGGTTACTTGCTTGCTTGCTTGCTTGCTTGCT 115239

Qy         464 TGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 523
               ||| || | ||||| | |   |  |||| |  || | |   |  |||| |  |||| |
Db      115240 TGCTTGCTTGCTTGCTTGCTTGTTTGCTTGATTGATTGATTGCTTGCTTGCTTGCTTGCG 115299

Qy         524 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTA 583
               |  |||| |  |||| |   |  |||| |  |||| |   |  | || |  |||| |
Db      115300 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGGT 115359

Qy         584 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 643
                  |||| |  |||| |   |  | || |  |||| |   |  |||| |  |||| |
Db      115360 TACTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 115419

Qy         644 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCT 703
               | | || |  |||| |   |  |||| |  |||| |   |  ||| |||| |||||||
Db      115420 TAGTTACTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 115479

Qy         704 TGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAAC 763
               |||||||||||| |   |  |  ||  |||  |   |    | |     |||  || |
Db      115480 TGCTTGCTTGCTTGCTTGTTCATTTTGCTGAGAAATGTTTAGGTTAAGGGCTAAGATAGA 115539

Qy         764 AGAGTAGGGCTGCCG 778
               |   ||| |||| |
Db      115540 TGTTCAGGTCTGCTG 115554


RESULT 5
AC097912/c
LOCUS       AC097912              218921 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-154L5, *** SEQUENCING IN PROGRESS
            ***.
ACCESSION   AC097912
VERSION     AC097912.5  GI:30520994
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 218921)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C.,, Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
```

```
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 218921)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 218921)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:22855232.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GFOX
                Center clone name: CH230-154L5
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 202489 bases at least Q40
                Consensus quality: 204424 bases at least Q30
                Consensus quality: 205792 bases at least Q20
                Estimated insert size: 214534; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *         1   218921: contig of 218921 bp in length.
FEATURES            Location/Qualifiers
     source         1. .218921
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-154L5"
     misc_feature   1. .1273
                    /note="wgs_contig"
BASE COUNT    57277 a  42734 c  45435 g  61021 t  12454 others
ORIGIN

  Query Match             24.1%;  Score 240.8;  DB 2;  Length 218921;
  Best Local Similarity   63.5%;  Pred. No. 2.5e-46;
  Matches  385;  Conservative    0;  Mismatches  217;  Indels    4;  Gaps    1;

Qy        130 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCG 189
              | ||  | | |||    ||||  |||||||| |      | | |   |||||| |||| |
Db      89107 CATGTCTTCAGAGTTTTCAGTTTGCCTGCTTACCTGCTTGCCTGCTTGCCTGCCTGCTTG 89048

Qy        190 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGG 249
              |  || ||| |||      | | |   | ||||| || |||| |   | ||||| ||| |
Db      89047 CCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCTTGCCTGCTTGCTTG 88988

Qy        250 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 309
              |||||||||||||| ||  || ||| |||||||  ||| ||||||||||| |||| ||  |
Db      88987 CCTGCTTGCTTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTG 88928

Qy        310 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 369
              | ||| |||||||  ||| ||||||  |||||||  ||| ||||||  |||||||  ||| |
Db      88927 CTTGCCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTGCTTG 88868

Qy        370 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 429
              |||||||||  |||| ||  ||||||||||  |||| ||  ||||||||||      | | |
Db      88867 CCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTGCCTGCTTGCT----TGCCTG 88812

Qy        430 TAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTG 489
              | |||||||| ||||      | | |   |||||||| | || ||    |   | |||
Db      88811 CTTGCCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTGCCTGCTTGCCTGCCTGCTTGCCTG 88752

Qy        490 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 549
              |  | || || |       |||  | |||| |   |  |||  |  |||| |   |  ||
Db      88751 CCTGCCTACTTGCCTACTTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCTTGCCTGCTTG 88692

Qy        550 CTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTC 609
              |  |  |||  |   | || || |  |||  |   | ||||| | ||||  |   | ||
Db      88691 CCTGCTTGCCTGCTTGCCTGCTTGCTTGCCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTA 88632

Qy        610 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 669
```

```
          || | ||||  |   | ||||| | ||||| |   | ||||| | ||||| |   |  ||
Db'    88631 CTTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTGCTTG 88572

Qy       670 CTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTG 729
             || | ||||  |   |  |||||||| |||| |||| || |||| ||    |    ||||
Db     88571 CTTGCCTGCCTGCCTGCTTGCCTGCTTGCTTCCTTGTTTCCTTGATATCTTGCTTCCTTG 88512

Qy       730 CCTGCT 735
             | || |
Db     88511 CTTGGT 88506


RESULT 6
AF210726
LOCUS       AF210726               130733 bp    DNA     linear   VRL 27-MAR-2000
DEFINITION  Macaca mulatta rhadinovirus 26-95 long unique region L-DNA,
            complete sequence.
ACCESSION   AF210726
VERSION     AF210726.1  GI:7329990
KEYWORDS    .
SOURCE      Macaca mulatta rhadinovirus 26-95
  ORGANISM  Macaca mulatta rhadinovirus 26-95
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Rhadinovirus.
REFERENCE   1  (bases 1 to 130733)
  AUTHORS   Alexander,L., Denekamp,L., Knapp,A., Auerbach,M.R., Damania,B. and
            Desrosiers,R.C.
  TITLE     The primary sequence of rhesus monkey rhadinovirus isolate 26-95:
            sequence similarities to Kaposi's sarcoma-associated herpesvirus
            and rhesus monkey rhadinovirus isolate 17577
  JOURNAL   J. Virol. 74 (7), 3388-3398 (2000)
  MEDLINE   20173730
   PUBMED   10708456
REFERENCE   2  (bases 1 to 130733)
  AUTHORS   Alexander,L., Denekamp,L.M., Knapp,A., Auerbach,M., Czajak,S.,
            Damania,B. and Desrosiers,R.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-DEC-1999) Microbiology, New England Regional Primate
            Research Center, One Pinehill Dr, Southborough, MA 01772, USA
FEATURES             Location/Qualifiers
     source          1. .130733
                     /organism="Macaca mulatta rhadinovirus 26-95"
                     /mol_type="genomic DNA"
                     /isolate="Macaca mulatta rhadinovirus isolate 26-95"
                     /db_xref="taxon:119193"
     gene            513. .1784
                     /gene="R1"
     CDS             513. .1784
                     /gene="R1"
                     /codon_start=1
                     /product="ORFR1"
                     /protein_id="AAF59980.1"
                     /db_xref="GI:7329991"
                     /translation="MFVLVLFMLLQPVSVELLPAKLTSVPTWCPPHPGDTYLLTCRGT
                     STARDQRSTQWFRNNTLMRGSNFYGRLVSVTPNATISDRYACQTKTTTRSNNIDFRVS
                     SSRLTLQERCSSYGYTYANNTRVLRCYSGGNVTLRNVVFHLNGTAVINGTTRNIHTFV
                     LTEKTGGTYFCSAFIGNEKFYSQTINVFFTSFTFKPTNDIPNESHFNKTGQIQQTASV
                     QHPENYVVFSVPVFSIGVLTGIAISLIMCWLFTIRCNENSESSTNSYASQTSYIQPSH
                     NQRSNTNECSRHTYRNAHQEESIEELPNQHTSETDSCCQLVLLEVKNVAYDGPQENTI
                     NEVMEQYDDVVVENIEQTSYEDNVEHMDYSDTINPNFNYYSGLILEEVDEVFYNELEN
                     QYHGLILENLDHNEYNHLNELNMIEQYDWLE"
     gene            complement(1852. .2418)
                     /gene="ORF02"
     CDS             complement(1852. .2418)
                     /gene="ORF02"
                     /codon_start=1
                     /product="DHFR"
                     /protein_id="AAF59981.1"
                     /db_xref="GI:7329992"
                     /translation="MDIAVNCIVAVDEQLGIGKNGTMPWPYLRNEMMYFQKMTSTPSV
                     VGEKNVVIMGKRTWFSIPEKKRPLVNRINIILSRELREPPHGAHFLARTLDDAFNFYR
                     QYKLKEQLNTVWVIGGKSVYESVLNYKCPLKLYITRIMESFDCDVFFPSINFTEYTML
                     SELPGKDTNFEENGIKYKFQVYEKNFNK"
     gene            2595. .3782
                     /gene="ORF04"
     CDS             2595. .3782
                     /gene="ORF04"
                     /codon_start=1
                     /product="complement binding protein"
                     /protein_id="AAF59982.1"
                     /db_xref="GI:7329993"
                     /translation="MFWIPITLCFICQVILVDSKDGENVTCLKPPFDRFMVKTANQNE
                     NYAVGTRVELICRPGFYKLQANVYVECLSNGTWTTPNAECRRKRCSNPEDILNGEVII
                     TDSDNAFKFGSNITYKCNTGYLLLGATVRTCLLKYDSNLVDWQPAAPTCEIEKCKKQP
                     DIENGKYYPVQEFYNYLETITFTCNKDFSLIGNTTTTCMTNGTWSSPVPKCQQITCSA
                     PNIDHGTLLVGSSRVYKHGQSVTIGCEDGFTLNGHKMCTCEYSLWNPPLPTCVPINKT
                     DPTPSEVPSPGTKEQEMPTPENPKSHESETTTTETPKTQTHKSETPSPKVPNPETHKP
                     TTPEAGISKQTTTNRPSKAPSQNPPIAPPMSKWKRHVVLVLFASVASLLFVLAALYCC
                     FLK"
     gene            4213. .7611
                     /gene="ORF06"
     CDS             4213. .7611
                     /gene="ORF06"
                     /codon_start=1
                     /product="ssDBP"
                     /protein_id="AAF59983.1"
                     /db_xref="GI:7329994"
                     /translation="MASKGNAGQPLEDNQGSRAPIGACGYVYAYSKQDFPFAEASILG
                     NRPSGSGVFSLPILYGLTVEHEFPLTVKAAYKKVDTTTLAVKVTCFHREVIVFHNASL
                     FRPVFDGTGLNELCEEARALFGYTQFIEPGPPHSIWNPLECPQLPDKDEMFLGVVVTE
                     GFKERLWRGCLVPAVFQTQQVQIAGRQAFKVPLYDEDLFAPHGHRMPRFYHKDVSAYL
                     YDSLFTSIAQALRLKDVTAVIHATEKQFMQDHYKIAKIVQAKQFSTTLPKTTDGSSHM
                     IVDSVVAELALSYGCMFLECPQDACELLNYDSWPIFDGCDSPEARVNALERWSAEQAV
                     HVAGQLFAANSVLYLTKVQKQAPRGQKGDVNVYNSFFLQHGLGFLNEATIKENGSEAF
                     KGVPSNALDGSSFTPYHLAYAASFSPHLLAKLCYYMQFLQHHKSSTNQAFNMVHYVGT
                     AANSEMCTLCHGNTPATCLNTLFYRLKDRFPAVTTPQRRDPYVVTGTAGTFNDLEILG
                     NFASFRDREEDGNPADEHPKYTYWQLCQTVTEKLSAIGITEDHDNHVNLITNIQSFLR
                     VFKGIDSIVDGEVMKFVNSMIKNNFNFREHVKSVHHILQFCCNVYWQAPCAVFLNLYY
                     KSLLWIIQDICLPYCMIYEQDNPAMGILPSEWLKMHFQTLWTNFKAACLDRGVLTGCE
                     LKIVHRDMFCDFFDTDAGSNGLMAPFKMQVRIARAMMVVPKSIKIKNRIIFSNTAGSE
                     AVQSGFVKPTGTRDTYVVAGPYMKFLNSLHRALFPDTKTAALYLWHKISQTNKTPVLK
                     DVPDDELAELVSYVKTNSLAFEETNVLDVVPDSLMSYARIKLNGAILRACGQIQFYAT
                     TLHCLTPVLQTIDAEEYPHVLGSAAIATPVAYLAEIRGRTALTVQTTARQPVAATGRL
                     RPVITVPMVVNKYTGVNGNNNVFHCGNLGYFAGRGVDRNLWPESSPFKKTGVSAMLRK
                     RHVMMTPIIDRLIKRAAGQTISTFEAESVKRSVQALLEDKDNPNLLKSVILELIRHLG
                     KGCQDLSSEDVQYYLGDYCMLTDEVLFTLDNIAQSGVPWTIEDAGALIEDRQDADDLQ
                     FVDSDDIATASCQPPEEQLPTPSAGALLAGKKRKINALLSDLDL"
     gene            7636. .9696
                     /gene="ORF07"
     CDS             7636. .9696
                     /gene="ORF07"
                     /codon_start=1
                     /product="transport protein"
                     /protein_id="AAF59984.1"
                     /db_xref="GI:7329995"
                     /translation="MARELAALYAQLSALAVDLSLVIFADPRSIDGARILKTKTQIEN
                     LNRDLLPLLREQNSVETSSLSLEVEHLAKNIEDKLGELERSLRQRYSSREHFETLHLR
                     PECHYHSTVTFQFYGGGLIDVNMCLINDVELLCKRLGSVFYCIGANEALSGLDRVLTF
                     LSTLRGISPIPHPDLYVTSVPCVQCLREIELVPNQGSSLLAVLADRHCDHLCKKVRAE
                     PIHGLFETELSQLGLKVTKRSDATQHGVRSSADQLRESSLAAIQDHNIFKRVSASIME
                     LSNLIYWNAGQTGLQTGTENECSQMARLLTHEADMHEHRALITPKLSATHFYDCFRPD
                     PIESLFCGGLFNSIDDTINALSRDCSVTFFQQANYTNVMRKQNELFTRLNSILRQGSA
                     GSQKPATPSEPRTTTVAATAASDVIKDAQYRKEQYMKKVARDGFKKLTECLQTQGAVL
                     ANALCMRVWGGVAYGEASELVNHFLLRRRFVALPWEARCRSDQILFENSKYIKNSLYS
                     QRLSREHVEIITLQFYGLITGPLTRQSDLFPGPANVVLAQCFEAAGMLPHHKMLVSEM
                     IWPQIQPKDWIDQTFNRFYQLPEGDLNAVQKSAWCFIRELVLSVALYNRTWEKTLRIF
                     SLAREKLSISNLDVKGLTSGLYLTYEQDAPLVLISQNTGWIFKDLYALLYHHLQLSDG
                     HDDN"
     gene            9683. .12172
                     /gene="ORF08"
     CDS             9683. .12172
                     /gene="ORF08"
                     /codon_start=1
                     /product="ORF08"
                     /protein_id="AAF59985.1"
                     /db_xref="GI:7329996"
```

```
                 /translation="MMITNRTRRLLRAWVVIIAIGTAVGENVTTPKGATTTAKPTPGP
                 STPTPPENPPRAEAFKFRVCSASATGELFRFNLEKTCPGTEDKTHQEGILMVFKKNIV
                 PHIFKVRRYRKVATSVTVYRGWTETAVTGKQEVIRPVPQYEINHMDTTYQCFSSMRVN
                 VNGIVNTYTDRDFTNQTVFLQPVEGLTDNIQRYFSQPVLYTTPGWFPGIYRVRTTVNC
                 EIVDMIARSAEPYSYFVTALGDTVEVSPFCHNDSTCSVAEKTENGLGVRALTNYTIVD
                 FATRKPTTETRVFADSGEYTVSWKAEDPKSAVCALTLWKTFPRAIQTTHEASYHFVAN
                 DVTATFTSPLSEVANFTGTYSCLDEVIQKTLNDTIKKLSDTHVTNGSAQYYKTEGGLF
                 LLWQPLTPLSLVDEMRGLNGTTPAPPATTSTVSRVRRSVNTNEQATDNLAAPQLQFAY
                 DKLRASINKVLEELSRAWCREQVRDTYMWYELSKINPTSVMTAIYGRPVSAKFVGDAI
                 SVTDCVAVDQASVSIHKSLRTSTPGMCYSRPPVTFRFLNSTTLFKGQLGPRNEIILTD
                 NQVEACKETCEHYFIASNVTYYYKDYVFVKKINTSEISTLGTFIALNLSFIENIDFRV
                 IELYSRAEKKLSGSVFDIETMFREYNYYTQRLAGLREDLDNTIDLNRDRLARDLSEIV
                 ADLGDVGRTVVNVASSVITLFGSIVSGFINFIKSPFGGMLMILVIVAVVLIVFALNRR
                 TNAIAQAPIRMIYPDIDKMQPSGGKVDQEQIKNILAGMHQLQQEERRRLDEQQRSAPS
                 LFRRASDGLKRRFRGYKPLENEEAQEYEMSK"
gene             12290. .15334
                 /gene="ORF09"
CDS              12290. .15334
                 /gene="ORF09"
                 /codon_start=1
                 /product="polymerase"
                 /protein_id="AAF59986.1"
                 /db_xref="GI:7329997"
                 /translation="MDFFNPYLGPRGPRPHSHRGTDAPAPAGAGAVQPPPDVCRLIPA
                 CLRTPGAGGMIPVTIPFPPTYFENGARGDVLLANERSMWTARDRKPVAPDPQDQSITF
                 HAYDVVETTYAADRCAEVPSRFQTDIIPSGTVLKLLGRTEDGTSVCVNVFRQQVYFYA
                 KVPAGINVTHILQQALKNTAGRAACGFSTRRVNKRILKTYDVAEHPVTEITLSSGSML
                 STLSDRLVACGCEVFESNVDAVRRFVLDHGFTTFGWYSCARATPRLAARDARTALEFD
                 CSWEDLSVQADRSDWPPYRIVAFDIECTGEAGFPCATRDGDAVIQISCVFYTTREGAP
                 NPPNILFSVGTCDPIPDTDVLEFPSEYDMLVSFFAMIRDFEVDFLTGYNISNFDLPYL
                 ITRASQVYNLRLNEYTKIKTGSIFEVHEPRGGGGGFMRSVSKIKIAGIVPIDMYQVCR
                 EKLSLSDYKLDTVARQCLGGKKEDVSYKDIPPLFRSGPGGRAKVGSYCVMDSVLVMDL
                 LKMFMIHVEISEIAKLAKIQARRVLTDGQQLRVFSCLLEAAARENFILPVPTPEGQGG
                 YQGATVINPIPGFYDEPVLVVDFASLYPSIIQAHNLCYSTMIHGRDLHLHPNLTPDDY
                 ETFVLSGGPVHFVKKHKRESLLGRLLTVWLEKRRAIRRTLAACDDPSLKTILDKQQLA
                 IKVTCNAVYGFTGVASGLLPCINIAETVTLRGRTMLEMSKSYVEALTTEDLRTRLGRE
                 VTARHGARFRVVYGDTDSLFIACDGYSAEAVSAFCDDLAARITADLFPPPIKLEAEKT
                 FKCLLLLTKKRYIGVLLNDKMVMKGVDLIRKTACKFVQERCRAILDLVLHDPEVKAAA
                 RLLCKRPPHAVYEEGLPAGFIKIVEVLNASYLDLRNSVVPIEQLTFSTELSRPVCDYK
                 TTNLPHLAVYQKLASRCEELPQVHDRIPYVFVDAPGSLKSDLAEHPDYVRQHQIPVAV
                 DLYFDKLVHGAANILQCLFGNNADTTVAILYNFLNVPYKLFS"
gene             15429. .16679
                 /gene="ORF10"
CDS              15429. .16679
                 /gene="ORF10"
                 /codon_start=1
                 /product="ORF10"
                 /protein_id="AAF59987.1"
                 /db_xref="GI:7329998"
                 /translation="MLVNELSVVLGDWEVTFHRGRFSFVNLTRLQTFKGHGGYARVRL
                 PFSLDQLLHQHFAFGLVTRLKELPPFSDCVALIAPLDSGGDADAARVAPGFVLDSSRP
                 LTVWVNASGRHTIRFCLLFLKPIDLERAVTYVFGENGGARSEGTPKPTCATESLPGGP
                 LRVSGEASQTSPHSFVAYFPTANSVACLSLLRLQVRPFSDDAAHRDARISPKYVTFSN
                 SGGNVCKASVHTLSPSRCKTAQMEIIYAPGDPNAEIVLGQSGPVLPTHTGGRVLGVYA
                 DAEKTIQPGSSAEVRVQLIFQQGAAARGDLAFLVTGVAPEPLFVVTPALLLSGCTTHL
                 RLFNPNGTPTTIKRDTLVAAAAPCPVVRLSSADDAPRDLVASPDTGALSINAFTIPVG
                 FPGVVSAECHVSLRDNGVHERMNH"
gene             16688. .17917
                 /gene="ORF11"
CDS              16688. .17917
                 /gene="ORF11"
                 /codon_start=1
                 /product="ORF11"
                 /protein_id="AAF59988.1"
                 /db_xref="GI:7329999"
                 /translation="MGTPVRFFRGEWQTSSLVDNGTPRYSSLVWAATIHDGYLTLVNR
                 SELCVTERSPCLPACPSIGRLVGKRFPGFAFASATLGDRGTRTVFYAFGHRDNPLDIV
                 PAVVERADRELVLRVHAPQTTRVSRYGLKVFVAIVTVVRPPGVFLHFPQDRVPIALTD

Query Match          23.9%;  Score 239.2;  DB 14;  Length 130733;
Best Local Similarity  88.1%;  Pred. No. 6.1e-46;
Matches  273;  Conservative    0;  Mismatches   33;  Indels    4;  Gaps    1;

Qy         69 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 128
              | |||| ||| |||||||||||||||||||      |   |   || ||||||||||||
Db     130428 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 130487

Qy        129 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 188
              ||||||||||      |   |  ||||||||||      |   |   || |||||||||
Db     130488 GCCTGCTTGCT----TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 130543

Qy        189 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 248
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     130544 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 130603

Qy        249 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 308
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     130604 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 130663

Qy        309 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 368
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     130664 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 130723

Qy        369 GCCTGCTTGC 378
              ||||||||||
Db     130724 GCCTGCTTGC 130733


RESULT 7
AC130110/c
LOCUS       AC130110               229535 bp    DNA     linear   HTG 22-SEP-2002
DEFINITION  Rattus norvegicus clone CH230-55M4, *** SEQUENCING IN PROGRESS ***,
            2 unordered pieces.
ACCESSION   AC130110
VERSION     AC130110.2  GI:23121680
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 229535)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
```

```
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 229535)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 229535)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 18, 2002 this sequence version replaced gi:22138333.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequening reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/).  As a result, the
            sequence may extend beyond the ends of the clone and there may be
            contigs that consist entirely of whole genome shotgun sequence
            reads.  Both end sequences and whole genome shotgun sequence only
            contigs will be indicated in the feature table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GJJP
                Center clone name: CH230-55M4
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 189188 bases at least Q40
                Consensus quality: 193629 bases at least Q30
                Consensus quality: 196703 bases at least Q20
                Estimated insert size: 207604; sum-of-contigs estimation
                Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   192206: contig of 192206 bp in length
            *  192207   192306: gap of unknown length
            *  192307   229535: contig of 37229 bp in length.
FEATURES            Location/Qualifiers
     source         1. .229535
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-55M4"
     misc_feature   48048. .49205
                    /note="wgs_contig"
     misc_feature   49707. .51011
                    /note="wgs_contig"
     misc_feature   complement(189833. .190293)
                    /note="clone_boundary
                    clone_end:T7
                    site:EcoRI
                    end_sequence:BH266273"
BASE COUNT    52992 a  49371 c  47396 g  49177 t  30599 others
ORIGIN

  Query Match             23.4%;  Score 234.4;  DB 2;  Length 229535;
  Best Local Similarity   58.2%;  Pred. No. 7.8e-45;
  Matches  412;  Conservative    0;  Mismatches  296;  Indels    0;  Gaps    0;

Qy          40 CGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCT 99
               ||  ||| || |||   |       |   || ||| ||  |||| | | || || ||| |
Db      211464 CGGTTGCTTGGTTGCTTGGTTGATTGCTTGGTTGCTTGGTTGCTTGGTTGCTTGGTTGGT 211405

Qy         100 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCT 159
                 |   |   || || ||| | | | |  |||||| |  |  |   |   || || |
Db      211404 TGGTTGCTTGGTTGCTTGGTTGTTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGT 211345

Qy         160 TGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGC 219
               ||||     |   |   || || ||| | | | |  || |||||  |  |   |   ||
Db      211344 TGCTTGCTTGGTTGCTTGCTTGGTTGGTTGTTTGGTTGGTTGCTTGGTTGGTTGCTTGGT 211285

Qy         220 TGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 279
               || | | ||| | |   || || | ||| |  || |||||||  | | | || || |||| |
Db      211284 TGGTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGT 211225

Qy         280 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 339
               |||| | | || || |||||||  | ||| || || ||| ||| | ||| |  || |||||
Db      211224 TGCTTGGTTGCTTGGTTGCTTGGTTGCTTGCTTGGTTGGTTGGTTGCTTGGTTGGTTGCT 211165

Qy         340 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 399
               |||| | | || |||||| ||||| | | || |||||| ||| | | | |  || ||| |
Db      211164 TGCTTGGTTGCTTGCTTGGTTGCTTGGTTGCTTGCTTGGTTGGTTGGTTGGTTGGTTGGT 211105

Qy         400 TGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTA 459
               ||||   | |  ||||||||      |   |  || || |||||||||  |  |   |
Db      211104 TGCTTAGTTGGTTGCTTGCTTGCTTGCTTGGTTGGTTGCTTGCTTGGTTGCTTGGT 211045

Qy         460 CGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 519
               |  ||  || | ||||| | |       |||| || || | |   || || | || || |
Db      211044 TGTTTGGTTGGTTGCTTGGTTGCTTCATTGCTTGGTTGGTTGGTTGGTTGGTTGGTTGGT 210985

Qy         520 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCT 579
               |   |  |||| |  |||| |   |  ||||  | || | |   || | || || || |
Db      210984 TGGTTGCTTGCTTGCTTGCTTGCTTGCTTAGTTGGTTGCTTGGTTGCTTGGTTGGT 210925

Qy         580 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 639
               |   || || | || || | |   | | || || |||| |   || |||| || ||||
Db      210924 TGGTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTGGTTGCTTGGTTGGTTGCTTGGTTGCT 210865

Qy         640 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCT 699
               |   |  |||| || |||| |   |  |||| || |||| |   |  ||| |||| |||
Db      210864 TGGTTGCTTGCTTGGTTGCTTGGTTGCTTGCTTGGTTGCTTGCTTGCTTGCTTGCTTGCT 210805

Qy         700 TGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCT 747
               |||||| |||||||||| |   |   | |||| || |     || || |
Db      210804 TGCTTGGTTGCTTGCTTGGTTGCTTGGTTGCTTGGTTGCTTGGTTGGT 210757


RESULT 8
AC131862/c
LOCUS       AC131862              245853 bp    DNA     linear   HTG 10-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-10F20, *** SEQUENCING IN PROGRESS
            ***.
ACCESSION   AC131862
VERSION     AC131862.3  GI:23674277
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
```

```
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 245853)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 245853)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 245853)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 10, 2002 this sequence version replaced gi:22539029.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GDSM
                Center clone name: CH230-10F20
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 227608 bases at least Q40
                Consensus quality: 230076 bases at least Q30
                Consensus quality: 231849 bases at least Q20
                Estimated insert size: 235882; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   245853: contig of 245853 bp in length.
FEATURES            Location/Qualifiers
     source         1. .245853
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-10F20"
     misc_feature   1. .1017
                    /note="wgs_contig"
     misc_feature   1068. .2461
                    /note="wgs_contig"
     misc_feature   2512. .3853
                    /note="wgs_contig"
     misc_feature   6786. .8519
                    /note="wgs_contig"
BASE COUNT    68006 a  49515 c  49070 g  66927 t  12335 others
ORIGIN

  Query Match             22.6%;  Score 226.2;  DB 2;  Length 245853;
  Best Local Similarity   64.1%;  Pred. No. 6.5e-43;
  Matches  357;  Conservative    0;  Mismatches  198;  Indels    2;  Gaps    1;

Qy        181 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCT 240
              ||||||| ||| || ||| ||||     |   |   |  |||| || |||| |   | ||
Db     175594 GCTTGCTAGCTTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTGCCTGCTTGCTTGCCT 175535

Qy        241 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG--CTTGC 298
              ||| ||| |||||||||| |||| ||| || ||| ||| |||| ||| || ||  |||||
Db     175534 GCTTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCCTGCCTGCTTGCTTGCTTGCTCTTGC 175475

Qy        299 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 358
               |||  ||  |||||| |||||||  ||  || ||||||||||||| ||| ||||||||||
Db     175474 CTGCCTGCCTGCCTGCCTGCTTGCCTGCCTGCTTGCTTGCTTGCTTGCCTGCTTGC 175415

Qy        359 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 418
              |||  ||  |||||||||||||| ||| || |||||||||||| ||| |||||| |||
Db     175414 CTGCCTGCCTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGCCTGC 175355
```

```
Qy        419 TGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGA 478
              |   |   | ||||||||||||||      |   |   |||| |||| | |||||
Db     175354 CTGCCTGCTTGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCCTACCTGCTTGCTTGC 175295

Qy        479 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 538
              | |   | ||||  | ||||  |   | |||||  |  ||||  |   |  | |  |  |||
Db     175294 TTGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCTTGCTTGCTTGCTTACCTGCTTGC 175235

Qy        539 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 598
              | |   |  |||| |  |||| |   |  | |  | ||||  |   | ||||  | ||||
Db     175234 TTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGC 175175

Qy        599 TAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 658
                |  || |  | ||||  |   | |||||  | |||||  |   |  ||||  | ||||
Db     175174 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCTCGCTTGCTCGCCTGC 175115

Qy        659 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGT 718
              | |   |  |||| |  |||| |   |  ||| |||| |||||||  ||| |||  ||| |
Db     175114 TTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTCGCTCGCTCGCTCGC 175055

Qy        719 GGGGCCGCTTGCCTGCT 735
              |  |||| ||  |||
Db     175054 TCGCTCGCTCGCTCGCT 175038


RESULT 9
AC106141
LOCUS       AC106141               305166 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-229F12, *** SEQUENCING IN PROGRESS
            ***, 17 unordered pieces.
ACCESSION   AC106141
VERSION     AC106141.4  GI:30579567
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 305166)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 305166)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 305166)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:23664597.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GJGS
                Center clone name: CH230-229F12
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 278355 bases at least Q40
                Consensus quality: 283506 bases at least Q30
                Consensus quality: 287188 bases at least Q20
                Estimated insert size: 285546; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3937: contig of 3937 bp in length
            *     3938     4037: gap of unknown length
            *     4038     8398: contig of 4361 bp in length
            *     8399     8498: gap of unknown length
```

```
            *     8499    11821: contig of 3323 bp in length
            *    11822    11921: gap of unknown length
            *    11922    27313: contig of 15392 bp in length
            *    27314    27413: gap of unknown length
            *    27414    31822: contig of 4409 bp in length
            *    31823    31922: gap of unknown length
            *    31923   272126: contig of 240204 bp in length
            *   272127   272226: gap of unknown length
            *   272227   275425: contig of 3199 bp in length
            *   275426   275525: gap of unknown length
            *   275526   282534: contig of 7009 bp in length
            *   282535   282634: gap of unknown length
            *   282635   283878: contig of 1244 bp in length
            *   283879   283978: gap of unknown length
            *   283979   285225: contig of 1247 bp in length
            *   285226   285325: gap of unknown length
            *   285326   286582: contig of 1257 bp in length
            *   286583   286682: gap of unknown length
            *   286683   288078: contig of 1396 bp in length
            *   288079   288178: gap of unknown length
            *   288179   289887: contig of 1709 bp in length
            *   289888   289987: gap of unknown length
            *   289988   291764: contig of 1777 bp in length
            *   291765   291864: gap of unknown length
            *   291865   295767: contig of 3903 bp in length
            *   295768   295867: gap of unknown length
            *   295868   299271: contig of 3404 bp in length
            *   299272   299371: gap of unknown length
            *   299372   305166: contig of 5795 bp in length.
FEATURES             Location/Qualifiers
     source          1. .305166
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-229F12"
     misc_feature    1. .1008
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    2758. .3559
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:RWBOD30TVB"
     misc_feature    23557. .25155
                     /note="wgs_contig"
     misc_feature    31923. .33139
                     /note="wgs_contig"
     misc_feature    237160. .237846
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:RWBOD30TJB"
     misc_feature    271133. .272126
                     /note="wgs_end_extension
                     clone_end:T7"
BASE COUNT    88552 a  60606 c  60203 g  80778 t  15027 others
ORIGIN

  Query Match             22.6%;  Score 226.2;  DB 2;  Length 305166;
  Best Local Similarity   64.1%;  Pred. No. 6.4e-43;
  Matches  357;  Conservative    0;  Mismatches  198;  Indels    2;  Gaps    1;

Qy        181 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCT 240
              ||||||| ||| || ||| ||||     |   |   |  |||| || |||| |   | ||
Db     247491 GCTTGCTAGCTTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTGCCTGCTTGCTTGCCT 247550

Qy        241 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG--CTTGC 298
              ||| ||| |||||||||| |||| ||| || ||| ||| |||| ||| || ||  |||||
Db     247551 GCTTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCCTGCCTGCTTGCTTGCTTGCTCTTGC 247610

Qy        299 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 358
              |||  ||  |||||| |||||||  ||  || ||||||||||| ||| ||||||||||
Db     247611 CTGCCTGCCTGCCTGCCTGCTTGCCTGCCTGCTTGCTTGCTTGCTTGCCTGCCTGCTTGC 247670

Qy        359 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 418
              |||  ||  |||||||||||||| ||| || |||||||||||| ||| ||||||  |||
Db     247671 CTGCCTGCCTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGCCTGC 247730

Qy        419 TGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGA 478
                  |   | | ||||||||||||||     |  |   |||| |||| | |||||
Db     247731 CTGCCTGCTTGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCCTACCTGCTTGCTTGC 247790

Qy        479 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 538
              | |   | ||||  | ||||  |   | |||||  |  ||||  |   |  | |  |  |||
Db     247791 TTGCTTGCCTGCCTGCCTGCCTGCCTGCTTGCTTGCTTGCTTGCTTACCTGCTTGC 247850

Qy        539 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 598
              | |   |  |||| |  |||| |   |  | |  | ||||  |   | ||||  | ||||
Db     247851 TTGCCTGCTTGCTTGCTTGCTTGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGC 247910

Qy        599 TAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 658
                |  | || |  | |||| |   | |||||  | |||||  |   |  ||||  | ||||
Db     247911 CTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCTCGCTTGCTCGCCTGC 247970

Qy        659 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGT 718
              | |   |  |||| |  |||| |   |  ||| |||| ||||||| ||| ||| ||| |
Db     247971 TTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTCGCTCGCTCGCTCGC 248030

Qy        719 GGGGCCGCTTGCCTGCT 735
              |  |||| ||  |||
Db     248031 TCGCTCGCTCGCTCGCT 248047


RESULT 10
AC137146/c
LOCUS       AC137146              208050 bp    DNA     linear   HTG 08-MAR-2003
DEFINITION  Mus musculus clone RP24-358I14, WORKING DRAFT SEQUENCE, 16
            unordered pieces.
ACCESSION   AC137146
VERSION     AC137146.2  GI:28882257
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 208050)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-358I14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 208050)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 208050)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
```

```
          *Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
           Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
           Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
           Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
           Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
           Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
           Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
           Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
           Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
           O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
           Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
           Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
           Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
           Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
           Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
           Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE     Direct Submission
JOURNAL   Submitted (08-MAR-2003) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Mar 8, 2003 this sequence version replaced gi:25048829.
          All repeats were identified using RepeatMasker:
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          -------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
          -------------- Project Information
              Center project name: L25975
              Center clone name: 358_I_14
          -------------- Summary Statistics
              Sequencing vector: Plasmid; n/a; 100% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.960731
              Consensus quality: 203377 bases at least Q40
              Consensus quality: 205277 bases at least Q30
              Consensus quality: 206093 bases at least Q20
              Insert size: 188000; agarose-fp
              Insert size: 206550; sum-of-contigs
              Quality coverage: 10.1 in Q20 bases; agarose-fp
              Quality coverage: 9.2 in Q20 bases; sum-of-contigs
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 16 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     1438: contig of 1438 bp in length
          *     1439     1538: gap of 100 bp
          *     1539     2410: contig of 872 bp in length
          *     2411     2510: gap of 100 bp
          *     2511     3868: contig of 1358 bp in length
          *     3869     3968: gap of 100 bp
          *     3969     5765: contig of 1797 bp in length
          *     5766     5865: gap of 100 bp
          *     5866     7076: contig of 1211 bp in length
          *     7077     7176: gap of 100 bp
          *     7177     8356: contig of 1180 bp in length
          *     8357     8456: gap of 100 bp
          *     8457     9860: contig of 1404 bp in length
          *     9861     9960: gap of 100 bp
          *     9961    11098: contig of 1138 bp in length
          *    11099    11198: gap of 100 bp
          *    11199    13131: contig of 1933 bp in length
          *    13132    13231: gap of 100 bp
          *    13232    15195: contig of 1964 bp in length
          *    15196    15295: gap of 100 bp
          *    15296    19695: contig of 4400 bp in length
          *    19696    19795: gap of 100 bp
          *    19796    24098: contig of 4303 bp in length
          *    24099    24198: gap of 100 bp
          *    24199    31817: contig of 7619 bp in length
          *    31818    31917: gap of 100 bp
          *    31918    44002: contig of 12085 bp in length
          *    44003    44102: gap of 100 bp
          *    44103   110199: contig of 66097 bp in length
          *   110200   110299: gap of 100 bp
          *   110300   208050: contig of 97751 bp in length.
FEATURES             Location/Qualifiers
     source          1. .208050
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-358I14"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .1438
                     /note="assembly_fragment"
     misc_feature    1539. .2410
                     /note="assembly_fragment"
     misc_feature    2511. .3868
                     /note="assembly_fragment"
     misc_feature    3969. .5765
                     /note="assembly_fragment"
     misc_feature    5866. .7076
                     /note="assembly_fragment"
     misc_feature    7177. .8356
                     /note="assembly_fragment"
     misc_feature    8457. .9860
                     /note="assembly_fragment"
     misc_feature    9961. .11098
                     /note="assembly_fragment"
     misc_feature    11199. .13131
                     /note="assembly_fragment"
     misc_feature    13232. .15195
                     /note="assembly_fragment"
     misc_feature    15296. .19695
                     /note="assembly_fragment"
     misc_feature    19796. .24098
                     /note="assembly_fragment"
     misc_feature    24199. .31817
                     /note="assembly_fragment"
     misc_feature    31918. .44002
                     /note="assembly_fragment"
     misc_feature    44103. .110199
                     /note="assembly_fragment"
     misc_feature    110300. .208050
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    58437 a  49871 c  45960 g  52282 t   1500 others
ORIGIN

  Query Match             22.5%;  Score 225.2;  DB 2;  Length 208050;
  Best Local Similarity   57.5%;  Pred. No. 1.1e-42;
  Matches  404;  Conservative    0;  Mismatches  298;  Indels    0;  Gaps    0;

Qy          45 GCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGG 104
               | |||  ||   | |  |  |   | ||||||||||||  ||  ||||||  |||
Db      161276 GTGTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 161217

Qy          105 GACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTG 164
               | | |   | |||| |||  ||  ||||||  |||       | | |   ||||||  |||
Db      161216 GCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 161157

Qy          165 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTG 224
                 | | |   |||||  |||  ||  | ||||  |||       | | |    | ||||
Db      161156 GCCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTGCCTGCCTGCCTGTCTGCCTGCCTGCCT 161097

Qy          225 GCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 284
               ||  |||  ||  | |||   ||  ||||||  |||  ||   ||  ||||||  |||  ||
Db      161096 GCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCT 161037
```

```
Qy        285 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 344
              ||  |||||| ||| |||  ||  |||||| ||| |||  ||  |||||| ||| |||
Db     161036 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 160977

Qy        345 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 404
              |  |||||| ||| |||  |  |||||| ||| |||  |  |||||| ||| |||
Db     160976 GTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCT 160917

Qy        405 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCT 464
              |  |||||| ||        | | |  | |||| ||| |||       | | |   | ||
Db     160916 GTCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCT 160857

Qy        465 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 524
              |||||   || ||   ||  | ||||  | ||||  ||  | ||||  | ||||  ||
Db     160856 GCCTGTCTGCCTGCCTGTCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGTCT 160797

Qy        525 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 584
              | ||||  | ||||  |   | ||||  | ||||  ||  | || |  | ||||  |
Db     160796 GCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGCCT 160737

Qy        585 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 644
              | ||||  | |||   |   | || |  | |||   |   | ||||  | |||   |
Db     160736 GCCTGCCTGTCTGTCTGCCTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGTCTGCCT 160677

Qy        645 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTT 704
              | ||||  | ||||  ||  | ||||  | ||||  |   | ||||||||  || ||| |
Db     160676 GCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCT 160617

Qy        705 GCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGC 746
              || ||| |||  |   | | || |||||||      | ||||
Db     160616 GCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGC 160575
```

RESULT 11
AC112675/c
LOCUS AC112675 228919 bp DNA linear HTG 21-OCT-2002
DEFINITION Mus musculus clone RP23-138J12, WORKING DRAFT SEQUENCE, 7 ordered pieces.
ACCESSION AC112675
VERSION AC112675.4 GI:24182124
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 228919)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP23-138J12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 228919)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (22-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 228919)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (21-OCT-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Oct 21, 2002 this sequence version replaced gi:21070725.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L22426
Center clone name: 138_J_12
---------- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 227366 bases at least Q40
Consensus quality: 227820 bases at least Q30
Consensus quality: 228087 bases at least Q20
Insert size: 230000; agarose-fp
Insert size: 228319; sum-of-contigs
Quality coverage: 12.8 in Q20 bases; agarose-fp
Quality coverage: 12.9 in Q20 bases; sum-of-contigs
----------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 48936: contig of 48936 bp in length
* 48937 49036: gap of 100 bp
* 49037 50559: contig of 1523 bp in length
* 50560 50659: gap of 100 bp
* 50660 56278: contig of 5619 bp in length
* 56279 56378: gap of 100 bp
* 56379 90153: contig of 33775 bp in length
* 90154 90253: gap of 100 bp
* 90254 114936: contig of 24683 bp in length
* 114937 115036: gap of 100 bp
* 115037 207735: contig of 92699 bp in length
* 207736 207835: gap of 100 bp
* 207836 228919: contig of 21084 bp in length.
FEATURES Location/Qualifiers
source 1. .228919
/organism="Mus musculus"

```
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-138J12"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .48936
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    49037. .50559
                     /note="assembly_fragment"
     misc_feature    50660. .56278
                     /note="assembly_fragment"
     misc_feature    56379. .90153
                     /note="assembly_fragment"
     misc_feature    90254. .114936
                     /note="assembly_fragment"
     misc_feature    115037. .207735
                     /note="assembly_fragment"
     misc_feature    207836. .228919
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    64052 a  55854 c  51189 g  57224 t    600 others
ORIGIN

  Query Match             22.5%;  Score 225.2;  DB 2;  Length 228919;
  Best Local Similarity   57.5%;  Pred. No. 1.1e-42;
  Matches  404;  Conservative    0;  Mismatches  298;  Indels    0;  Gaps    0;

Qy         45 GCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGG 104
              | |||  ||   | |  |  |    | ||||||||||||  ||  |||||| |||
Db        801 GTGTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 742

Qy        105 GACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTG 164
              | | |   | ||||  |||  ||  ||||||  |||       | | |   ||||||  |||
Db        741 GCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 682

Qy        165 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTG 224
                | | |   |||||  |||  ||  | ||||  |||       | | |   | ||||
Db        681 GCCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTGCCTGCCTGCCTGTCTGCCTGCCTGCCT 622

Qy        225 GCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 284
              ||  |||  ||  | |||   ||  ||||||  |||  ||   ||  ||||||  |||  ||
Db        621 GCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCT 562

Qy        285 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 344
              ||  ||||||  |||  |||   ||  ||||||  |||  |||   ||  ||||||  |||  |||
Db        561 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 502

Qy        345 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 404
              |   ||||||  |||  |||   |   ||||||  |||  |||   |   ||||||  |||  |||
Db        501 GTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCT 442

Qy        405 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCT 464
              |   ||||||  ||       | | |   | ||||  |||  |||       | | |   | ||
Db        441 GTCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCT 382

Qy        465 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 524
              |||||   ||  ||   ||  | ||||  | ||||  ||  | ||||  | ||||  ||
Db        381 GCCTGTCTGCCTGCCTGTCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGTCT 322

Qy        525 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 584
              | ||||  | ||||  |   | ||||  | ||||  ||  | ||  | |  | ||||  |
Db        321 GCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGCCT 262

Qy        585 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 644
              | ||||  | |||   |   | ||  |  | |||   |   | ||||  | |||   |
Db        261 GCCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGTCTGCCT 202

Qy        645 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTT 704
              | ||||  | ||||  ||  | ||||  | ||||  |   | ||||||||  || ||| |
Db        201 GCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCT 142

Qy        705 GCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGC 746
                ||  ||| |||  |   | | ||  |||||||      | ||||
Db        141 GCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGC 100


RESULT 12
AC115409
LOCUS       AC115409              220464 bp    DNA     linear   HTG 13-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-134D4, WORKING DRAFT SEQUENCE.
ACCESSION   AC115409
VERSION     AC115409.4  GI:24941929
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 220464)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Haml,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 220464)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2002) Human Genome Sequencing Center, Department
```

```
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 220464)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 13, 2002 this sequence version replaced gi:22857077.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GOMT
                Center clone name: CH230-134D4
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 207904 bases at least Q40
                Consensus quality: 209884 bases at least Q30
                Consensus quality: 210779 bases at least Q20
                Estimated insert size: 213748; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   220464: contig of 220464 bp in length.
FEATURES            Location/Qualifiers
     source         1. .220464
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-134D4"
     misc_feature   1. .2329
                    /note="wgs_contig"
     misc_feature   219334. .220464
                    /note="wgs_contig"
BASE COUNT    60311 a  46976 c  47584 g  56550 t   9043 others
ORIGIN

  Query Match             21.9%;  Score 218.6;  DB 2;  Length 220464;
  Best Local Similarity   56.7%;  Pred. No. 4e-41;
  Matches  404;  Conservative    0;  Mismatches  309;  Indels    0;  Gaps    0;

Qy          44 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 103
               ||| ||| ||   | |       | ||||||||||   ||  |||||| |||
Db       81468 TGCCTGCCTGCCTGCCTGTCTCCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTC 81527

Qy         104 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 163
               | | ||  |||||| |||  ||  |||||  ||       | | ||  |||||  |||
Db       81528 TGCCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTGCCTGCCTGTCTGCCTGTCTGCC 81587

Qy         164 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 223
               | | |   ||||||  |||  ||  ||||||  |||      | | |   | ||||
Db       81588 TGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCC 81647

Qy         224 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 283
               |  |||  |   | ||||  ||  | ||||  |||  |||  ||  | ||||  |||  ||
Db       81648 TGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGTC 81707

Qy         284 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 343
               ||  ||||||  |||  ||   ||  ||||||  |||  |||   ||  ||||||  ||  |||
Db       81708 TGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGTCTGCC 81767

Qy         344 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403
               ||  ||||||  ||  |||  ||  ||||||  |||  |||   ||  ||||||  ||  ||
Db       81768 TGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGTC 81827

Qy         404 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCC 463
               ||  |||||  |||       | | |   | ||||  |||  |||       | |  ||  |||
Db       81828 TGCCTGCCTGTCTGCCTGTCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCC 81887

Qy         464 TGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 523
               ||||||   ||  ||   ||  | ||||  | ||||  ||  | ||||  | ||||  ||
Db       81888 TGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTC 81947

Qy         524 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTA 583
               | ||||  | |||   |   | |||   | ||||  |   | ||  |  | |||   ||
Db       81948 TGCCTGCCTGCCTGTCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGTC 82007

Qy         584 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 643
               | ||||  | ||||  ||  | ||  | ||||  |   | ||||  | |||   |
Db       82008 TGCCTGCCTGTCTGCCTGTCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCC 82067

Qy         644 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCT 703
               | ||||  | ||||  ||  | ||||  | |||   |     ||||||||  |  |||
Db       82068 TGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGTCTGCCTACCTGCCTGCCTGTCTGCC 82127

Qy         704 TGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTG 756
               ||| ||  |||  |   | | || |||||||      | ||||    | ||||
Db       82128 TGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCTG 82180


RESULT 13
AC107913
LOCUS       AC107913              128953 bp    DNA     linear   PRI 28-AUG-2002
DEFINITION  Homo sapiens chromosome 17, clone CTD-3051C7, complete sequence.
ACCESSION   AC107913
VERSION     AC107913.7  GI:22532638
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 128953)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone CTD-3051C7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 128953)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
```

```
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 128953)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 128953)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 28, 2002 this sequence version replaced gi:22297439.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L23044
                Center clone name: 3051_C_7
            --------------
FEATURES             Location/Qualifiers
     source          1. .128953
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="CTD-3051C7"
                     /clone_lib="CITD2 Human BAC"
     repeat_region   386. .461
                     /rpt_family="CT-rich"
     repeat_region   complement(1888. .1975)
                     /rpt_family="tRNA-Arg-AGA"
     repeat_region   2505. .2586
                     /rpt_family="tRNA-Leu-CTA"
     repeat_region   complement(3077. .3148)
                     /rpt_family="tRNA-Gln-CAG"
     repeat_region   3246. .3523
                     /rpt_family="MER33"
     repeat_region   complement(3670. .3745)
                     /rpt_family="tRNA-Lys-AAA"
     repeat_region   5228. .5356
                     /rpt_family="MIR3"
     repeat_region   complement(5464. .5758)
                     /rpt_family="AluSq"
     repeat_region   5760. .5812
                     /rpt_family="MER91"
     repeat_region   complement(6843. .7005)
                     /rpt_family="L2"
     repeat_region   8462. .8650
                     /rpt_family="C-rich"
     repeat_region   complement(8858. .9155)
                     /rpt_family="AluSp"
     repeat_region   complement(9226. .9529)
                     /rpt_family="AluSx"
     repeat_region   complement(9582. .9876)
                     /rpt_family="L2"
     repeat_region   9965. .10118
                     /rpt_family="FAM"
     repeat_region   11131. .11308
                     /rpt_family="MIR"
     repeat_region   12520. .12540
                     /rpt_family="GC_rich"
     repeat_region   13116. .13238
                     /rpt_family="C-rich"
     repeat_region   complement(14519. .14584)
                     /rpt_family="MIR"
     repeat_region   14825. .15001
                     /rpt_family="(TTTC)n"
     repeat_region   15002. .15390
                     /rpt_family="(CTTG)n"
     repeat_region   complement(15391. .15680)
                     /rpt_family="AluY"
     repeat_region   complement(15684. .15965)
                     /rpt_family="AluSq"
     repeat_region   complement(16437. .16572)
                     /rpt_family="AluSc"
     repeat_region   complement(16656. .16785)
                     /rpt_family="FLAM_A"
     repeat_region   17009. .17100
                     /rpt_family="L2"
     repeat_region   complement(17263. .17542)
                     /rpt_family="AluJo"
     repeat_region   complement(17569. .18117)
                     /rpt_family="L2"
     repeat_region   complement(18137. .18354)
                     /rpt_family="L2"
     repeat_region   19136. .19250
                     /rpt_family="(TTGGGG)n"
     repeat_region   complement(20134. .20296)
                     /rpt_family="L2"
     repeat_region   complement(20313. .20372)
                     /rpt_family="L2"
     repeat_region   complement(20407. .20614)
                     /rpt_family="L2"
     repeat_region   complement(20717. .20993)
                     /rpt_family="AluSg"
     repeat_region   21268. .21571
                     /rpt_family="AluJb"
```

```
     repeat_region   21879. .22182
                     /rpt_family="AluSq"
     repeat_region   complement(22225. .22349)
                     /rpt_family="MIR"
     repeat_region   22422. .22595
                     /rpt_family="FAM"
     repeat_region   complement(22673. .22850)
                     /rpt_family="MLT1H1"
     repeat_region   complement(22902. .23026)
                     /rpt_family="MLT1H1"
     repeat_region   23051. .23106
                     /rpt_family="CT-rich"
     repeat_region   complement(23157. .23416)
                     /rpt_family="AluJo"
     repeat_region   complement(23521. .23824)
                     /rpt_family="AluSx"
     repeat_region   23863. .24015
                     /rpt_family="MIR"
     repeat_region   24050. .24358
                     /rpt_family="AluJo"
     repeat_region   24377. .24669
                     /rpt_family="L1M4"
     repeat_region   24670. .24981
                     /rpt_family="AluSx"
     repeat_region   24982. .25142

  Query Match           21.8%;  Score 218.2;  DB 9;  Length 128953;
  Best Local Similarity   69.1%;  Pred. No. 5.1e-41;
  Matches  298;  Conservative    0;  Mismatches  133;  Indels    0;  Gaps    0;

Qy         72 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 131
              ||| ||| |||| ||| || ||||||||     |   |   || ||||||| ||| ||
Db      14961 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 15020

Qy        132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191
              ||||||||     |   |   || ||||||||     |   |   || |||||||| |||
Db      15021 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 15080

Qy        192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251
               || ||||||||     |   |   |  |||| ||||||| |   |  |||| ||| ||
Db      15081 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 15140

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
              |||||||||||| ||| || ||||||||||||| ||| || |||||||||||| ||| ||
Db      15141 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 15200

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
              |||||||||||| ||| || |||||||||||| ||| || |||||||||||| ||| ||
Db      15201 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 15260

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTA 431
              |||||||||||| ||| || |||||||||||| ||| || ||||||||     |   |
Db      15261 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 15320

Qy        432 GGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCT 491
               |  ||||||||||||     |   |   || ||| || | ||||| | |   |  ||||
Db      15321 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 15380

Qy        492 GGGCTGCTAGT 502
               |  |||| ||
Db      15381 TGCTTGCTTGT 15391


RESULT 14
AC141883
LOCUS       AC141883               238215 bp    DNA     linear   HTG 20-MAR-2003
DEFINITION  Mus musculus chromosome UNK clone RP24-128P19, WORKING DRAFT
            SEQUENCE, 31 unordered pieces.
ACCESSION   AC141883
VERSION     AC141883.2  GI:29126550
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 238215)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 238215)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 238215)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Mar 20, 2003 this sequence version replaced gi:29124284.

            ---------------- Genome Center ----------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information -------------------
            Center project name: M_BB0128P19
            ------------------------------------------------------
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; %
            Sequencing vector: plasmid; %
            Chemistry: Dye-primer ET; % of reads
            Chemistry: Dye-terminator Big Dye; % of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 221212 bases at least Q40
            Consensus quality: 224001 bases at least Q30
            Consensus quality: 225617 bases at least Q20
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 31 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1212: contig of 1212 bp in length
            *     1213     1312: gap of unknown length
            *     1313     2808: contig of 1496 bp in length
            *     2809     2908: gap of unknown length
            *     2909     4509: contig of 1601 bp in length
            *     4510     4609: gap of unknown length
            *     4610     6126: contig of 1517 bp in length
            *     6127     6226: gap of unknown length
            *     6227     7673: contig of 1447 bp in length
            *     7674     7773: gap of unknown length
            *     7774     8968: contig of 1195 bp in length
            *     8969     9068: gap of unknown length
            *     9069    10583: contig of 1515 bp in length
            *    10584    10683: gap of unknown length
            *    10684    11778: contig of 1095 bp in length
            *    11779    11878: gap of unknown length
            *    11879    13464: contig of 1586 bp in length
            *    13465    13564: gap of unknown length
            *    13565    14971: contig of 1407 bp in length
            *    14972    15071: gap of unknown length
            *    15072    16342: contig of 1271 bp in length
            *    16343    16442: gap of unknown length
            *    16443    17795: contig of 1353 bp in length
            *    17796    17895: gap of unknown length
            *    17896    19947: contig of 2052 bp in length
            *    19948    20047: gap of unknown length
            *    20048    21207: contig of 1160 bp in length
            *    21208    21307: gap of unknown length
            *    21308    23316: contig of 2009 bp in length
```

```
            *    23317    23416: gap of unknown length
            *    23417    24878: contig of 1462 bp in length
            *    24879    24978: gap of unknown length
            *    24979    26136: contig of 1158 bp in length
            *    26137    26236: gap of unknown length
            *    26237    27558: contig of 1322 bp in length
            *    27559    27658: gap of unknown length
            *    27659    29825: contig of 2167 bp in length
            *    29826    29925: gap of unknown length
            *    29926    31576: contig of 1651 bp in length
            *    31577    31676: gap of unknown length
            *    31677    33353: contig of 1677 bp in length
            *    33354    33453: gap of unknown length
            *    33454    35537: contig of 2084 bp in length
            *    35538    35637: gap of unknown length
            *    35638    37238: contig of 1601 bp in length
            *    37239    37338: gap of unknown length
            *    37339    38481: contig of 1143 bp in length
            *    38482    38581: gap of unknown length
            *    38582    40522: contig of 1941 bp in length
            *    40523    40622: gap of unknown length
            *    40623    45987: contig of 5365 bp in length
            *    45988    46087: gap of unknown length
            *    46088    70413: contig of 24326 bp in length
            *    70414    70513: gap of unknown length
            *    70514   101243: contig of 30730 bp in length
            *   101244   101343: gap of unknown length
            *   101344   136512: contig of 35169 bp in length
            *   136513   136612: gap of unknown length
            *   136613   183419: contig of 46807 bp in length
            *   183420   183519: gap of unknown length
            *   183520   238215: contig of 54696 bp in length.
FEATURES             Location/Qualifiers
     source          1. .238215
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP24-128P19"
     misc_feature    1. .1212
                     /note="assembly_name:Contig13"
     misc_feature    1313. .2808
                     /note="assembly_name:Contig14"
     misc_feature    2909. .4509
                     /note="assembly_name:Contig16"
     misc_feature    4610. .6126
                     /note="assembly_name:Contig17"
     misc_feature    6227. .7673
                     /note="assembly_name:Contig18"
     misc_feature    7774. .8968
                     /note="assembly_name:Contig19"
     misc_feature    9069. .10583
                     /note="assembly_name:Contig20"
     misc_feature    10684. .11778
                     /note="assembly_name:Contig26"
     misc_feature    11879. .13464
                     /note="assembly_name:Contig27"
     misc_feature    13565. .14971
                     /note="assembly_name:Contig28"
     misc_feature    15072. .16342
                     /note="assembly_name:Contig30"
     misc_feature    16443. .17795
                     /note="assembly_name:Contig31"
     misc_feature    17896. .19947
                     /note="assembly_name:Contig33"
     misc_feature    20048. .21207
                     /note="assembly_name:Contig34"
     misc_feature    21308. .23316
                     /note="assembly_name:Contig35"
     misc_feature    23417. .24878
                     /note="assembly_name:Contig37"
     misc_feature    24979. .26136
                     /note="assembly_name:Contig38"
     misc_feature    26237. .27558
                     /note="assembly_name:Contig40"
     misc_feature    27659. .29825
                     /note="assembly_name:Contig41"
     misc_feature    29926. .31576
                     /note="assembly_name:Contig42"
     misc_feature    31677. .33353
                     /note="assembly_name:Contig43"
     misc_feature    33454. .35537
                     /note="assembly_name:Contig44"
     misc_feature    35638. .37238
                     /note="assembly_name:Contig45"
     misc_feature    37339. .38481
                     /note="assembly_name:Contig46"
     misc_feature    38582. .40522
                     /note="assembly_name:Contig47"
     misc_feature    40623. .45987
                     /note="assembly_name:Contig48"
     misc_feature    46088. .70413
                     /note="assembly_name:Contig49"
     misc_feature    70514. .101243
                     /note="assembly_name:Contig50"
     misc_feature    101344. .136512
                     /note="assembly_name:Contig51"
     misc_feature    136613. .183419
                     /note="assembly_name:Contig52"
     misc_feature    183520. .238215
                     /note="assembly_name:Contig53"
BASE COUNT    62298 a  54994 c  55470 g  62353 t   3100 others
ORIGIN

  Query Match             21.3%;  Score 213;  DB 2;  Length 238215;
  Best Local Similarity   63.2%;  Pred. No. 8.1e-40;
  Matches  327;  Conservative    0;  Mismatches  190;  Indels    0;  Gaps    0;

Qy         65 GTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 124
              ||   ||||  ||  |||| ||  || || |||||     | | |   |||||||||| |
Db     102060 GTGTTGCTGTTTGTTTGCTTGCCTGCTTGTTTGCTTGCCTGCCTGCTTGCCTGCTTGTTT 102119

Qy        125 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 184
              ||| |||||||||| |     | | |   ||||||||||       |   |   ||||||||
Db     102120 GCTTGCCTGCTTGTTTGCTTGTCTGCCTGCCTGCTTGCCTGCCTGTTTGCTTGCCTGCTT 102179

Qy        185 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTC 244
              | | ||| |||||||||| |     | | |   |  |||  |||||| |   | ||||
Db     102180 GTTTGCTTGCCTGCTTGTTTGCTTGCCTGCTTGCTTGCCTGCTTGCTTGCCTGACTGCCT 102239

Qy        245 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 304
              ||  || ||||||||||||| ||  || ||| ||||||  ||| |||||||||| ||||
Db     102240 GCCTGCTTGCTTGCTTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTT 102299

Qy        305 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 364
              ||  || ||| |||||||  ||| |||||| ||| |||  ||  |||||||||| ||||
Db     102300 GCCTGCTTGCCTGCTTGCCTGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTT 102359

Qy        365 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 424
              | | || ||| |||||||  ||| |||||||||| |||| ||  ||||||  ||||
Db     102360 GTTTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCCTGCCTGCTTGCCT 102419

Qy        425 GACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAG 484
              |   ||  |  ||| |||||||      |   |   ||||||  || | |||||   |
Db     102420 GCTTGTTTGCTTGCCTGCTTGCCTGCTTGTTTGCTTGCCTGCTTGCTTGCTTGCCTGCTT 102479

Qy        485 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 544
              | ||||  | ||||| |   | ||||  |  |||  |   | |||||  | ||||  |
Db     102480 GCCTGCCTGCCTGCTTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCCTGCCT 102539

Qy        545 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAG 581
              |  |||  | ||||  |   |  | || |  |||| |
Db     102540 GCTTGCCTGCCTGCCTGCCTGCTTGCTTGCTTGCTTG 102576


RESULT 15
```

```
AC132225
LOCUS       AC132225               219713 bp    DNA     linear   HTG 03-SEP-2002
DEFINITION  Mus musculus chromosome UNK clone RP24-558H21, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC132225
VERSION     AC132225.1  GI:22657606
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 219713)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 219713)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-SEP-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT
            ---------------- Genome Center ----------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information -------------------
            Center project name: M_BB0558H21
            ------------------------------------------------------
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 216515 bases at least Q40
            Consensus quality: 216969 bases at least Q30
            Consensus quality: 217277 bases at least Q20
            Insert size: 208000; agarose-fp
            Insert size: 218913; sum-of-contigs
            Quality coverage: 14.28 in Q20 bases; agarose-fp
            Quality coverage: 10.95 in Q20 bases; sum-of-contigs
            ------------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    11635: contig of 11635 bp in length
            *    11636    11735: gap of unknown length
            *    11736    24748: contig of 13013 bp in length
            *    24749    24848: gap of unknown length
            *    24849    41932: contig of 17084 bp in length
            *    41933    42032: gap of unknown length
            *    42033    66899: contig of 24867 bp in length
            *    66900    66999: gap of unknown length
            *    67000    95249: contig of 28250 bp in length
            *    95250    95349: gap of unknown length
            *    95350   143163: contig of 47814 bp in length
            *   143164   143263: gap of unknown length
            *   143264   202040: contig of 58777 bp in length
            *   202041   202140: gap of unknown length
            *   202141   209252: contig of 7112 bp in length
            *   209253   209352: gap of unknown length
            *   209353   219713: contig of 10361 bp in length.
FEATURES             Location/Qualifiers
     source          1. .219713
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP24-558H21"
     misc_feature    1. .11635
                     /note="assembly_name:Contig10"
     misc_feature    11736. .24748
                     /note="assembly_name:Contig11"
     misc_feature    24849. .41932
                     /note="assembly_name:Contig12"
     misc_feature    42033. .66899
                     /note="assembly_name:Contig13"
     misc_feature    67000. .95249
                     /note="assembly_name:Contig14"
     misc_feature    95350. .143163
                     /note="assembly_name:Contig15"
     misc_feature    143264. .202040
                     /note="assembly_name:Contig16"
     misc_feature    202141. .209252
                     /note="assembly_name:Contig8"
     misc_feature    209353. .219713
                     /note="assembly_name:Contig9"
BASE COUNT    58319 a  48550 c  50247 g  61747 t    850 others
ORIGIN

  Query Match             20.3%;  Score 202.6;  DB 2;  Length 219713;
  Best Local Similarity   55.5%;  Pred. No. 2.2e-37;
  Matches  391;  Conservative    0;  Mismatches  314;  Indels    0;  Gaps    0;

Qy         42 TGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGA 101
              ||| ||| ||   | |  |   |  | ||||||||||||  ||    |||| |||
Db      82470 TCTGCCTGCCTGCCTGCCTGCCTCTCTGCCTGCCTGCCTGCCTGCCTCTCTGCCTGCCTG 82529

Qy        102 GGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTG 161
                |    | |||||| |||  ||  ||||||  |||       | |  |  |||||| ||
Db      82530 CCTGCATATCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTATCTGCCTGCCTG 82589

Qy        162 CTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTG 221
              |      | |  |  ||||||  |||  ||  ||||||  |||         | |   | |||
Db      82590 CCTGCATGCCTATCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCTCTGCCTGCCTG 82649

Qy        222 CTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 281
              |     |||  |   | ||||  ||  ||||||  |||  |||  ||  ||||||  |   ||
Db      82650 CCTCTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTATCTG 82709

Qy        282 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 341
              |  ||  ||||||  ||| |    ||  || |||  |||  |||       |||||| ||| ||
Db      82710 CCTGCCTGCCTGCCTGCCTCTCTGCCTGCTTGCCTGCCTGCCTCTCTGCCTGCCTGCCTG 82769

Qy        342 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 401
              |      ||||||  |||  ||  ||    |||| |||  |||  ||  |||||| ||| |
Db      82770 CCTATCTGCCTGCCTGCCTGCCTGCCTATCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTA 82829

Qy        402 CTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACG 461
                ||  ||||||  |||      | | |     |||| |||  |||       | |   |
Db      82830 TCTGCCTGCCTGCCTGCCTGCCTGCCTCTCTGCCTGCCTGCCTGCCTATCTGCCTG 82889

Qy        462 CCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAG 521
              ||||||||   || ||   |  | ||    | ||||  |   | ||||  |  ||||  |
Db      82890 CCTGCCTGCCTGCCTGCCTGCCTCTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTG 82949

Qy        522 TAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAG 581
              | ||||  | ||||  |   | ||||  | ||||  |   | || |  | ||||  |
Db      82950 CCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTATCTGCCTGCCTGCCTGCCTG 83009

Qy        582 TAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAG 641
              |  | ||||  | ||||  |   | || |  | ||||  |     ||||  | ||||
Db      83010 TCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTATCTGCCTGCCTGCCTC 83069

Qy        642 TAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTG 701
              |  | |||||  | ||||   |  | ||||  | ||||   |  | ||||||||  || |
Db      83070 TCTGCCTGCTTGCCTGCCTCTCTGCCTGCCTGCCTGCCTATCTGCCTGCCTGCCTGCCTC 83129

Qy        702 CTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGC 746
                ||| ||| ||   |   | | || |||||||      | ||||
```

```
Db      83130 TCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGC 83174


RESULT 16
AC073686
LOCUS       AC073686               229010 bp    DNA     linear   HTG 29-JUN-2000
DEFINITION  Mus musculus clone RP23-137O3, WORKING DRAFT SEQUENCE, 24 unordered
            pieces.
ACCESSION   AC073686
VERSION     AC073686.1  GI:8810303
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 229010)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 229010)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA.
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 1789819
            Center clone name: RPCI-23_137O3
            -------
            Summary Statistics
            Consensus quality: 210402 bases at least Q40
            Consensus quality: 220249 bases at least Q30
            Consensus quality: 222247 bases at least Q20
            Estimated insert size: 231000; agarose-fp estimation
            Estimated insert size: 226710; sum-of-contigs estimation
            Quality coverage:  8.2 in Q20 bases; agarose-fp estimation
            Quality coverage:  8.35 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 24 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1213: contig of 1213 bp in length
            *     1214     1313: gap of unknown length
            *     1314     2782: contig of 1469 bp in length
            *     2783     2882: gap of unknown length
            *     2883     3987: contig of 1105 bp in length
            *     3988     4087: gap of unknown length
            *     4088     5464: contig of 1377 bp in length
            *     5465     5564: gap of unknown length
            *     5565     6708: contig of 1144 bp in length
            *     6709     6808: gap of unknown length
            *     6809     9727: contig of 2919 bp in length
            *     9728     9827: gap of unknown length
            *     9828    11170: contig of 1343 bp in length
            *    11171    11270: gap of unknown length
            *    11271    14915: contig of 3645 bp in length
            *    14916    15015: gap of unknown length
            *    15016    17525: contig of 2510 bp in length
            *    17526    17625: gap of unknown length
            *    17626    19558: contig of 1933 bp in length
            *    19559    19658: gap of unknown length
            *    19659    24103: contig of 4445 bp in length
            *    24104    24203: gap of unknown length
            *    24204    31169: contig of 6966 bp in length
            *    31170    31269: gap of unknown length
            *    31270    36961: contig of 5692 bp in length
            *    36962    37061: gap of unknown length
            *    37062    44441: contig of 7380 bp in length
            *    44442    44541: gap of unknown length
            *    44542    52366: contig of 7825 bp in length
            *    52367    52466: gap of unknown length
            *    52467    61496: contig of 9030 bp in length
            *    61497    61596: gap of unknown length
            *    61597    77798: contig of 16202 bp in length
            *    77799    77898: gap of unknown length
            *    77899    88928: contig of 11030 bp in length
            *    88929    89028: gap of unknown length
            *    89029   102867: contig of 13839 bp in length
            *   102868   102967: gap of unknown length
            *   102968   119649: contig of 16682 bp in length
            *   119650   119749: gap of unknown length
            *   119750   138547: contig of 18798 bp in length
            *   138548   138647: gap of unknown length
            *   138648   157565: contig of 18918 bp in length
            *   157566   157665: gap of unknown length
            *   157666   184818: contig of 27153 bp in length
            *   184819   184918: gap of unknown length
            *   184919   229010: contig of 44092 bp in length.
FEATURES             Location/Qualifiers
     source          1. .229010
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-137O3"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT    63071 a  50891 c  49844 g  62897 t   2307 others
ORIGIN

 Query Match             20.3%;  Score 202.6;  DB 2;  Length 229010;
  Best Local Similarity   63.2%;  Pred. No. 2.2e-37;
  Matches  327;  Conservative    0;  Mismatches  189;  Indels    1;  Gaps    1;

Qy         65 GTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 124
              ||   ||||  ||  |||| ||  || || |||||     | | |   |||||||| |
Db      37954 GTGTTGCTGTTTGTTTGCTTGCCTGCTTGTTTGCTTGCCTGCCTGCTTGCCTGCTTGTTT 38013

Qy        125 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 184
              ||| |||||||||| |     | | |   ||||||||||      |   |  ||||||||
Db      38014 GCTTGCCTGCTTGTTTGCTTGTCTGCCTGCCTGCTTGCCTGCCTGTTTGCTTGCCTGCTT 38073

Qy        185 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTC 244
              | | ||| |||||||| |     | | |   |  |||  ||||||| |   | ||||
Db      38074 GTTTGCTTGCCTGCTTGTTTGCTTGCCTGCTTGCTTGCCTGCTTGCTTGCCTGACTGCCT 38133

Qy        245 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 304
              ||  || |||||||||||| ||  || ||| |||||||  ||| ||||||||| ||||
Db      38134 GCCTGCTTGCTTGCTTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTT 38193

Qy        305 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 364
              ||  || ||| |||||||  ||| |||||| ||| |||  ||  |||||||||| ||||
Db      38194 GCCTGCTTGCCTGCTTGCCTGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTT 38253

Qy        365 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 424
              | | || ||| |||||||  ||| ||||||||| |||| ||  |||||| ||||
Db      38254 GTTTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCCTGCCTGCTTGCCT 38313

Qy        425 GACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAG 484
              |   |   | ||| |||||||      |  | |  |||||| || | ||||   |
Db      38314 GCTTGGTTGCTTGCCTGCTTGCCTGCTTGTTGCT-TGCCTGCTTGCTTGCTTGCCTGCTT 38372

Qy        485 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 544
              | ||||  | ||||| |   | ||||  |  |||  |   | ||||| | ||||  |
Db      38373 GCCTGCCTGCCTGCTTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCCTGCCT 38432

Qy        545 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAG 581
              |  |||  | ||||  |   |  | || ||  |||| |
Db      38433 GCTTGCCTGCCTGCCTGCCTGCTTGCTTGCTTGCTTG 38469
```

```
RESULT 17
AC098079
LOCUS       AC098079              276954 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-126O22, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC098079
VERSION     AC098079.6  GI:30522704
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 276954)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 276954)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 276954)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:23269184.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GGFU
                Center clone name: CH230-126O22
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 219903 bases at least Q40
                Consensus quality: 222552 bases at least Q30
                Consensus quality: 224170 bases at least Q20
                Estimated insert size: 232969; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   271976: contig of 271976 bp in length
            *   271977   272076: gap of unknown length
            *   272077   276954: contig of 4878 bp in length.
FEATURES             Location/Qualifiers
     source          1. .276954
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-126O22"
     misc_feature    1. .1105
                     /note="wgs_contig"
     misc_feature    35576. .37460
                     /note="wgs_contig"
BASE COUNT    71863 a  46380 c  44551 g  62022 t  52138 others
ORIGIN

  Query Match             20.2%;  Score 202;  DB 2;  Length 276954;
  Best Local Similarity   56.9%;  Pred. No. 3.1e-37;
  Matches  370;  Conservative    0;  Mismatches  280;  Indels    0;  Gaps    0;

Qy        125 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 184
              | |||     | | ||   ||   || |||  ||| ||||   ||   |  |||  |||
Db     215111 GGTGGTTGTTTGGTTGGTTGGCTGGTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTG 215170

Qy        185 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTC 244
              ||| ||||||  ||| ||||   ||    | ||   | || || ||| |  ||   |||
Db     215171 GCTGGCTGGCTGGCTGGCTGGCTGGCTGTTTGGTTGGTTGACTGGCTGGCTGGCTGGCTG 215230

Qy        245 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 304
```

```
          |||||| ||| ||| ||| |||||| ||| ||| ||| |||||| ||| ||| | |
Db     215231 GCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGTTG 215290

Qy        305 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 364
              | |||| | | | | | |||||  ||| ||| ||| |||||| ||| ||| |||
Db     215291 GTTGGCTGGTTGGTTGGTTGACTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTG 215350

Qy        365 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 424
              |||||| ||| ||| ||| |||||| ||| ||| ||| | |||  | | | ||   |
Db     215351 GCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGTTGGCTGGTTGGTTGGTTG 215410

Qy        425 GACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAG 484
              |   |  ||  ||| ||| |||    ||  |||  |   ||  | ||| | | |    |
Db     215411 GCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGTTGGTTGGCTGGTTGGTTGGTTGACTG 215470

Qy        485 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 544
              |  |||||  ||| |  ||  |||||  ||| |  ||  |||||  ||| |  |
Db     215471 GCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTG 215530

Qy        545 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 604
              |  |||||  |||  | ||     ||    ||| |  ||  |||||  ||| |  |
Db     215531 GCTGGCTGGCTGGCTGTTTGGTTGGTTGACTGGCTGGCTGGCTGGCTGGCTGGCTGGCTG 215590

Qy        605 GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 664
              |  ||||  ||| |  ||  |||||  |||  | ||  | ||    ||| |  |
Db     215591 GCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGTTTGGTTGGTTGACTGGCTGGCTG 215650

Qy        665 GGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCC 724
              |  |||||  ||| |  ||  ||  ||||||| ||| |||  | | | || ||
Db     215651 GCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGTTTGGTTGGTTGGCTG 215710

Qy        725 GCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGGGCT 774
              | ||  |  ||  |   |     | || |  |||  ||| | ||   ||
Db     215711 GTTTTGCCACTTTTCATTCCATACTCCATCCTGGAAGACATAATATTCCT 215760
```

RESULT 18
AC117991
LOCUS AC117991 271190 bp DNA linear HTG 13-MAR-2003
DEFINITION Mus musculus clone RP23-148J22, WORKING DRAFT SEQUENCE, 17 unordered pieces.
ACCESSION AC117991
VERSION AC117991.3 GI:28933908
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 271190)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP23-148J22
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 271190)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 271190)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouellei1,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (13-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 13, 2003 this sequence version replaced gi:28209782.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L15575
Center clone name: 148_J_22
---------- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 267377 bases at least Q40
Consensus quality: 268701 bases at least Q30
Consensus quality: 269200 bases at least Q20
Insert size: 236000; agarose-fp
Insert size: 269590; sum-of-contigs
Quality coverage: 10.8 in Q20 bases; agarose-fp
Quality coverage: 9.5 in Q20 bases; sum-of-contigs
----------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 710: contig of 710 bp in length
* 711 810: gap of 100 bp
* 811 5531: contig of 4721 bp in length
* 5532 5631: gap of 100 bp
* 5632 7825: contig of 2194 bp in length
* 7826 7925: gap of 100 bp
* 7926 13529: contig of 5604 bp in length
* 13530 13629: gap of 100 bp
* 13630 19264: contig of 5635 bp in length
* 19265 19364: gap of 100 bp
* 19365 25381: contig of 6017 bp in length
* 25382 25481: gap of 100 bp

```
            *    25482    38157: contig of 12676 bp in length
            *    38158    38257: gap of 100 bp
            *    38258    52292: contig of 14035 bp in length
            *    52293    52392: gap of 100 bp
            *    52393    70736: contig of 18344 bp in length
            *    70737    70836: gap of 100 bp
            *    70837    82153: contig of 11317 bp in length
            *    82154    82253: gap of 100 bp
            *    82254   101307: contig of 19054 bp in length
            *   101308   101407: gap of 100 bp
            *   101408   120674: contig of 19267 bp in length
            *   120675   120774: gap of 100 bp
            *   120775   149796: contig of 29022 bp in length
            *   149797   149896: gap of 100 bp
            *   149897   176329: contig of 26433 bp in length
            *   176330   176429: gap of 100 bp
            *   176430   202707: contig of 26278 bp in length
            *   202708   202807: gap of 100 bp
            *   202808   234746: contig of 31939 bp in length
            *   234747   234846: gap of 100 bp
            *   234847   271190: contig of 36344 bp in length.
FEATURES             Location/Qualifiers
     source          1. .271190
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-148J22"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .710
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    811. .5531
                     /note="assembly_fragment"
     misc_feature    5632. .7825
                     /note="assembly_fragment"
     misc_feature    7926. .13529
                     /note="assembly_fragment"
     misc_feature    13630. .19264
                     /note="assembly_fragment"
     misc_feature    19365. .25381
                     /note="assembly_fragment"
     misc_feature    25482. .38157
                     /note="assembly_fragment"
     misc_feature    38258. .52292
                     /note="assembly_fragment"
     misc_feature    52393. .70736
                     /note="assembly_fragment"
     misc_feature    70837. .82153
                     /note="assembly_fragment"
     misc_feature    82254. .101307
                     /note="assembly_fragment"
     misc_feature    101408. .120674
                     /note="assembly_fragment"
     misc_feature    120775. .149796
                     /note="assembly_fragment"
     misc_feature    149897. .176329
                     /note="assembly_fragment"
     misc_feature    176430. .202707
                     /note="assembly_fragment"
     misc_feature    202808. .234746
                     /note="assembly_fragment"
     misc_feature    234847. .271190
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    70565 a  63542 c  62426 g  73055 t   1602 others
ORIGIN

  Query Match          20.1%;  Score 201.2;  DB 2;  Length 271190;
  Best Local Similarity   56.9%;  Pred. No. 4.7e-37;
  Matches  388;  Conservative    0;  Mismatches  293;  Indels    1;  Gaps    1;
```

```
Qy         77 GCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTT 136
              ||  | | | ||||  ||| ||||   ||    |  ||   ||| | | | |||   | |
Db     186653 GCTAGTTAGTTGGCTGGCTGGCTGGCTGGTTGGCTGGTTGGCTGGTTGGTTGGTTGGTTG 186712

Qy        137 GCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 196
              ||||   ||   |  | |   || ||||   ||   || ||   ||| ||| ||||||
Db     186713 GCTGGCTGGCTGGCTGTCAGTCTGGCTGGTTGGCTGGTTGGTTAGCTGGCTGGCTGGCTG 186772

Qy        197 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 256
              ||| ||||   ||  ||| ||   | ||| | | | || ||   | | | ||||  |||
Db     186773 GCTGGCTGGTTGGTTAGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCTG 186832

Qy        257 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 316
              | | ||| |||||   ||| ||| |||  | | |   || | | | | ||||||  |||
Db     186833 GTTGGCTGGCTGGTTGGCTGGCTGGCTGTCAGTCTGTCTGGTTGGTTGGCTGGCTGGCTG 186892

Qy        317 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 376
              ||| ||| ||||||  ||| ||| ||| ||||||  | | ||| | | | ||||  | |
Db     186893 GCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGTTGGCTGGTTGGTTGGCTAGTTA 186952

Qy        377 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 436
              | | ||| ||||||  ||| ||| | | |||||   ||| | ||   ||   || ||
Db     186953 GTTGGCTGGCTGGCTGGCTGGCTGGTTGGCTGGTTGGCTGGTTGGTTGGTTGGCTG 187012

Qy        437 GCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCT 496
              ||| |||  |      |   |||  ||  |   | | ||| | | |  ||   |||||
Db     187013 GCTGGCTGTCAGTCTGTCTGGTTGGCTGGTTGGTTAGCTGGCTGGCTGGCTGGCTGGTTG 187072

Qy        497 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 556
              | ||||  || || | || || | |   || || | || ||   |  |||||  ||||
Db     187073 GTTAGT-TGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGGCTGGCTGGCT 187131

Qy        557 GCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCT 616
              | ||||  | ||  | ||||| | |   |||||  |||||   |   ||||   ||||
Db     187132 GTTAGTCTGTCTGGTTGGCTGGTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 187191

Qy        617 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 676
              |   |   || ||   | ||| | |   || || | || || | | |  |||||  ||||
Db     187192 GGCTGGCAGGTTGTCAGTCTGGTTGGCTGGTTGTTTGGGTGGTTGGTTGGCTGACTGGCT 187251

Qy        677 GCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTA 736
              |   |   || ||  || | | |||  || ||| ||| | |   ||  | ||| |||
Db     187252 GGCTGGCTGGTTGGTTGGTTGTTTGGGTGGTTGGTTGGTTGGTTGGTTGGCTGGCT 187311

Qy        737 CTAGGGCTGCTGTGCAGCTGGG 758
                  ||||| |   |  | | ||
Db     187312 GGCTGGCTGGTTGGTTGGTTGG 187333
```

```
RESULT 19
AC096116
LOCUS       AC096116               224055 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-10O16, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC096116
VERSION     AC096116.7  GI:30522727
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 224055)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
```

```
            'Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 224055)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 224055)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24817844.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GELV
                Center clone name: CH230-10O16
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 196024 bases at least Q40
                Consensus quality: 199448 bases at least Q30
                Consensus quality: 201887 bases at least Q20
                Estimated insert size: 204330; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   222891: contig of 222891 bp in length
            *  222892   222991: gap of unknown length
            *  222992   224055: contig of 1064 bp in length.
FEATURES             Location/Qualifiers
     source          1. .224055
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-10O16"
     misc_feature    1. .1326
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    2494. .3408
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH308507"
     misc_feature    53120. .54604
                     /note="wgs_contig"
     misc_feature    123391. .126252
                     /note="wgs_contig"
     misc_feature    221867. .222775
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH308510"
BASE COUNT    54935 a  43967 c  46140 g  58898 t  20115 others
ORIGIN

  Query Match           19.5%;  Score 195.4;  DB 2;  Length 224055;
  Best Local Similarity   55.5%;  Pred. No. 1.1e-35;
  Matches  376;  Conservative    0;  Mismatches  301;  Indels    0;  Gaps    0;

Qy         81 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 140
              ||| ||||||  | | ||||   ||   || ||   | | | | | ||||  ||| | ||
Db     154410 GCTGGCTGGCTGGTTGGCTGGTTGGCTGGTTGGTTGGTTGGTTGGCTGGCTGGTTG 154469

Qy        141 AGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTT 200
                ||   || ||   | | ||||   ||   |  |||  ||| ||| |||||     |
Db     154470 GTTGGTTGGTTGGTTGGTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGTTGATTG 154529

Qy        201 GCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTT 260
              | ||   ||   || ||   ||||||| | |||| ||   ||| ||||||  ||| | |
Db     154530 GTTGGTTGGTTGGTTGGCTGGCTGGCTGGTTAGTTGGTTGGCTGGCTGGCTGGCTGGTTG 154589

Qy        261 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 320
              | | ||||||  ||| ||| | | | |||   | | | | | | ||||||  ||| | |
Db     154590 GTTGGCTGGCTGGCTGGCTGGTTGGTTGGTTGGTTAGTTGGTTGGCTGGCTGGCTGGTTA 154649

Qy        321 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 380
              | | | ||||  ||| ||| ||| ||||||  | |   | | | | |||   | | |||
Db     154650 GTTGGTTGGCTGGCTGGCTGGCTGGCTGGTTGATTGGTTGGTTGGTTGGTTGGCTG 154709

Qy        381 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTT 440
```

```
Db     154710 GCTAGCTGGTTAGTTGGTTGGCTGGCTGGGCGGCTGGCTGGTTGGTTGGTTAGTTGGTTG 154769

Qy        441 GCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTA 500

Db     154770 GCTGGCTGGGTGGCTGGCTGGCTGGCTGGCTGGTTGGTTGGTTGGTTGGTTGGCTGGCTG 154829

Qy        501 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 560

Db     154830 GGTGGCTGGCTGGTTGATTGGTTGGTTGGTTGGCTGGATGGCTGGTTGGTTGGTTGGTTG 154889

Qy        561 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTA 620

Db     154890 GTTGGCTGGATGGCTGGTTAGTTGGTTGGCTGGCTGGGTGGCTGGCTGGCTGGCTGGCTG 154949

Qy        621 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 680

Db     154950 GTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCTAGCTGGTTAGTTGGTTGGCTGGCTG 155009

Qy        681 GTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAG 740

Db     155010 GGTGGCTGGCTGGTTGGTTGGTTAGTTGGTTGGCTGGCTGGGTGGCTGGCTGGCTGGCTG 155069

Qy        741 GGCTGCTGTGCAGCTGG 757

Db     155070 GCTGGCTGGTTGGTTGG 155086


RESULT 20
AC110455/c
LOCUS       AC110455               183863 bp    DNA     linear   HTG 15-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-276C17, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC110455
VERSION     AC110455.6  GI:25006850
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 183863)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 183863)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 183863)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 15, 2002 this sequence version replaced gi:22856759.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GRIZ
                Center clone name: CH230-276C17
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 169880 bases at least Q40
                Consensus quality: 170862 bases at least Q30
                Consensus quality: 171472 bases at least Q20
                Estimated insert size: 169890; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   181513: contig of 181513 bp in length
            *  181514   181613: gap of unknown length
```

```
            *   181614   182673: contig of 1060 bp in length
            *   182674   182773: gap of unknown length
            *   182774   183863: contig of 1090 bp in length.
FEATURES             Location/Qualifiers
     source          1. .183863
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-276C17"
     misc_feature    1. .1195
                     /note="wgs_contig"
     misc_feature    180129. .181513
                     /note="wgs_contig"
BASE COUNT    51113 a  33333 c  34320 g  52915 t  12182 others
ORIGIN

  Query Match             19.4%;  Score 194.4;  DB 2;  Length 183863;
  Best Local Similarity   58.4%;  Pred. No. 1.9e-35;
  Matches  339;  Conservative    0;  Mismatches  241;  Indels    0;  Gaps    0;

Qy        169 GACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTT 228
              | | ||  |||||| |||  ||  |||||| |||      | | ||  | ||||  || |
Db     169634 GCCTGTGTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTGTGCCTGCCTGCCT 169575

Qy        229 GCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 288
              |  |   | |||| ||  |||||  ||  |||  ||  |||||| ||| |||  ||
Db     169574 ACCTGCCTGCCTGCCTGCCAGCCTGTGTGACTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 169515

Qy        289 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 348
              |||||| ||| |||  ||  |||||| ||  |||  ||  |||||| ||| |||  ||
Db     169514 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGTGTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 169455

Qy        349 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 408
              |||||| ||| |||  ||  |||||| ||| |||  ||  |||||| ||| |||  ||
Db     169454 GCCTGCCTGCGTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 169395

Qy        409 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCT 468
              |||||| |||    | | | |   | |||| ||| |||      | | |   ||||||||
Db     169394 GCCTGCCTGCCTGTGTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 169335

Qy        469 GATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 528
              |   || ||  |   | |||   | ||||  |   | ||||  | | ||  ||  | ||
Db     169334 GCCTGCCTGCCTGCCTGCCTGTGTGCCTGCCTGCCTGCCTGCCTGCCAGCCTGTGTGCCT 169275

Qy        529 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCT 588
              ||  | ||||  |    ||||  | ||||  |   | ||    | ||||  |    ||
Db     169274 GCCTGCCTGCCTGCCTACCTGCCTGCCTGCCTGCCTGCCTGTGTGCCTGCCTGCCTACCT 169215

Qy        589 GCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 648
              ||  | ||||  |   | || |  | ||||  |   | ||||  |  |||  |   | ||
Db     169214 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTGTGCCTGCCTGCCT 169155

Qy        649 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTT 708
              ||  | ||||  |   | || |  |  |||  ||  | ||||||||  || ||| ||| |
Db     169154 GCCTGCCTGCCTGCCTGCCTCCCTGTGTGCCTGTGTGCCTGCCTGCCTGCCTGCCTGCCT 169095

Qy        709 GCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTG 748
              || |||  |   | | || ||||||||  | | |    |||
Db     169094 GCCTGCCTGCCTGCCTGCCTGCCTGCTGCCATGCTACCTG 169055


RESULT 21
AP000291
LOCUS       AP000291               27213 bp    DNA     linear   PRI 20-NOV-1999
DEFINITION  Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region,
            clone:Q64C8, complete sequence.
ACCESSION   AP000291
VERSION     AP000291.1  GI:4835660
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 27213)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE     Homo sapiens 27,213bp genomic DNA of 21q22.1
  JOURNAL   Published Only in DataBase (1999)
REFERENCE   2  (bases 1 to 27213)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     The sequence is a part of the data (ACCESSION No. AP000174 -
            AP000194).
            The sequencing project is supported by Japan Science Technology
            Corporation (JST) and The Institute of Physical and Chemical
            Research (RIKEN).
FEATURES             Location/Qualifiers
     source          1. .27213
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="Q64C8"
BASE COUNT     7998 a   5908 c   5753 g   7554 t
ORIGIN

  Query Match             19.3%;  Score 192.8;  DB 9;  Length 27213;
  Best Local Similarity   59.2%;  Pred. No. 5.2e-35;
  Matches  329;  Conservative    0;  Mismatches  227;  Indels    0;  Gaps    0;

Qy        180 TGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGC 239
              || ||| |   | |  |||||| |  | |   |   |  || | |  || | |   ||
Db     20372 TGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTTGGCTGGTTGCTTGGT 20431

Qy        240 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 299
              || | ||| |  |||||| ||| | | | |  |||||||||||| ||| || ||||||||
Db     20432 TGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 20491

Qy        300 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 359
              |||| ||| || ||  || ||| | | | |  |||||| ||| | | | || ||||||||
Db     20492 TGCTTGCTTGCTTGGCTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGCTTGCTTGCT 20551

Qy        360 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              || | ||| || |  ||| |||||  | | || |||||  ||| | ||| |  ||||||||
Db     20552 TGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCT 20611

Qy        420 GAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGAT 479
              |  |   |   || || |||||||  |  |   |   |  ||  ||  | | | ||| |
Db     20612 TCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGCT 20671

Qy        480 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 539
              |   || |||| || |||| |   || || | || || | ||  || |||| || || |
Db     20672 TGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGTTTGGTTGCTTGGTTGGT 20731

Qy        540 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 599
              |   || || | |  || | |   || |  | |  |||| |    | |||| |  || |
Db     20732 TGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTTGTTGCTTGCTTGGT 20791

Qy        600 AGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 659
              || || | || || | |   || || | || || | |   || || | |  | || |
Db     20792 TCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTCGTTGAT 20851

Qy        660 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTG 719
              |   || || | || |||| |   || || ||| |   ||| ||| ||| ||||| |
Db     20852 TGCTTGGTTGGTTGGTTGCTTGGTTGGTTGGCTGGTTCATTGGTTGGTTGGTTGCTTGGT 20911

Qy        720 GGGCCGCTTGCCTGCT 735
              |   | |||  || |
```

```
Db      20912 TGTTTGGTTGGTTGGT 20927


RESULT 22
AP000043
LOCUS       AP000043              100000 bp    DNA     linear   PRI 20-NOV-1999
DEFINITION  Homo sapiens genomic DNA, chromosome 21q22.1, segment 14/28,
            complete sequence.
ACCESSION   AP000043
VERSION     AP000043.1  GI:3132353
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA, chromosome 21q
  JOURNAL   Published Only in DataBase (1998)
REFERENCE   2  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAY-1998) Masahira Hattori, Kitasato University,
            Department of Science, JST Sequencing Laboratory; Kitasato 1-15-1,
            Sagamihara 228, Japan (E-mail:hattori@hgc.ims.u-tokyo.ac.jp,
            Tel:0427-78-9732, Fax:0427-78-9561)
COMMENT     This sequence is conducted by Kitasato University JST sequencing
            Laboratory as a JST sequencing team.
            Principal Investigator:Yoshiyuki Sakaki Ph.D.
            Phone: +81-3-5449-5622, Fax : +81-3-5449-5445,
            sakaki@hgc.ims.u-tokyo.ac.jp
            Sub-leader: Tadayoshi Shiba Ph.D., Masahira Hattori Ph.D. The
            sequence is submitted by:Human Genome Sequencing in ALIS project of
            JST
            Japan Science and Technology Corporation (JST)
            5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0028 Japan
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www-alis.tokyo.jst.go.jp/HGS/top. html)
            or send email to webmaster@www-alis.tokyo.jst.go.jp'.
FEATURES             Location/Qualifiers
     source          1. .100000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="T1866-f90A8f_7"
BASE COUNT    27784 a  22015 c  21661 g  28540 t
ORIGIN

  Query Match             19.3%;  Score 192.8;  DB 9;  Length 100000;
  Best Local Similarity   59.2%;  Pred. No. 4.7e-35;
  Matches  329;  Conservative    0;  Mismatches  227;  Indels    0;  Gaps    0;

Qy        180 TGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGC 239
              || ||| |   | |  |||||| |  |  |   |   |  || | |  || | |   ||
Db      12956 TGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTTGGCTGGTTGCTTGGT 13015

Qy        240 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 299
              || | ||| |  |||||| ||| | | | |  |||||||||||| ||| || ||||||||
Db      13016 TGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 13075

Qy        300 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 359
              |||| ||| || ||  || ||| | | | |  |||||| ||| | | | || ||||||||
Db      13076 TGCTTGCTTGCTTGGCTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGCTTGCTTGCT 13135

Qy        360 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              || | ||| || |  ||| |||| | | || |||||  ||| | ||| |  ||||||||
Db      13136 TGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCT 13195

Qy        420 GAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGAT 479
                |  |   |  || || |||||| |  |  |   |   |  ||  || | | ||| |
```

```
Db      13196 TCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCT 13255

Qy        480 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 539
              |   || |||| || |||| |   || || | || || | ||  || |||| || || |
Db      13256 TGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGTTTGGTTGCTTGGTTGGT 13315

Qy        540 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 599
              |   || || | |  || | |   || | |  |||| |    | |||| |  || |
Db      13316 TGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTTGTTGCTTGCTTGGT 13375

Qy        600 AGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 659
              || || | || || | |   || || | || || | |   || || |  | || |
Db      13376 TCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTCGTTGAT 13435

Qy        660 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTG 719
              |   || || | || |||| |   || || ||| |   ||| ||| ||| ||||| |
Db      13436 TGCTTGGTTGGTTGGTTGCTTGGTTGGTTGGCTGGTTCATTGGTTGGTTGGTTGCTTGGT 13495

Qy        720 GGGCCGCTTGCCTGCT 735
              |   | |||  || |
Db      13496 TGTTTGGTTGGTTGGT 13511


RESULT 23
AP000111
LOCUS       AP000111              100000 bp    DNA     linear   PRI 25-MAY-2002
DEFINITION  Homo sapiens genomic DNA of 21q22.1, GART and AML related,
            Q78C10-149C3 region, segment 14/20.
ACCESSION   AP000111
VERSION     AP000111.1  GI:4730845
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Homo sapiens 2,051,516bp genomic DNA of 21q22.1 GART and AML region
  JOURNAL   Published Only in Database (1999)
REFERENCE   2  (bases 1 to 100000)
  AUTHORS   Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-1999) Mika Hirakawa, Japan Science and Technology
            Corporation (JST), Advanced Databases Department; 5-3, Yonbancho,
            Chiyoda-ku, Tokyo 102-0081, Japan (E-mail:mika@tokyo.jst.go.jp,
            URL:http://www-alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
            Fax:81-3-5214-8470)
COMMENT     This sequence is conducted by Kitasato University JST sequencing
            Laboratory as a JST sequencing team.
            Principal Investigator:Yoshiyuki Sakaki Ph.D.
            Phone: +81-3-5449-5622, Fax : +81-3-5449-5445,
            sakaki@hgc.ims.u-tokyo.ac.jp
            Sub-leader: Tadayoshi Shiba Ph.D., Masahira Hattori Ph.D. The
            sequence is submitted by Human Genome Sequencing in ALIS project of
            JST.
            Japan Science and Technology Corporation (JST)
            5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081 Japan
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www-alis.tokyo.jst.go.jp/HGS/) or send
            email to webmaster@www-alis.tokyo.jst.go.jp.
FEATURES             Location/Qualifiers
     source          1. .100000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
     STS             66264. .66403
                     /note="A006D32;The location is between each flanking site
                     of PCR primers."
                     /db_xref="GDB:4571163"
     STS             66282. .66515
```

```
                /note="SHGC-52112;The location is between each flanking
                site of PCR primers."
                /db_xref="GDB:6464755"
     STS        92086. .92206
                /note="SHGC-52139;The location is between each flanking
                site of PCR primers."
                /db_xref="GDB:6464581"
     STS        92099. .92279
                /note="RH18027;The location is between each flanking site
                of PCR primers."
                /db_xref="GDB:4572813"
BASE COUNT    27992 a  22056 c  21621 g  28331 t
ORIGIN

  Query Match             19.3%;  Score 192.8;  DB 9;  Length 100000;
  Best Local Similarity   59.2%;  Pred. No. 4.7e-35;
  Matches  329;  Conservative    0;  Mismatches  227;  Indels    0;  Gaps    0;

Qy        180 TGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGC 239
              || ||| |   | |  |||||| |  |  |   |   |  || | |  || | |   ||
Db      18932 TGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTTGGCTGGTTGCTTGGT 18991

Qy        240 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 299
              || | ||| |  |||||| ||| | | | |  |||||||||||| ||| || ||||||||
Db      18992 TGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 19051

Qy        300 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 359
              |||| ||| || ||  || ||| | | | |  |||||| ||| | | | || ||||||||
Db      19052 TGCTTGCTTGCTTGGCTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGCTTGCTTGCT 19111

Qy        360 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              || | ||| || |  ||| ||||| | | || |||||  ||| | ||| |  ||||||||
Db      19112 TGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCT 19171

Qy        420 GAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGAT 479
                |  |   |   || || ||||||| |  |  |   |   |  ||  || | | ||| |
Db      19172 TCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGCT 19231

Qy        480 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 539
               |   || |||| || |||| |   || || | || || | |'  || |||| || || |
Db      19232 TGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGTTTGGTTGCTTGGTTGGT 19291

Qy        540 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 599
               |   || || | |  || | |   || |  | |  |||| |    | |||| |  || |
Db      19292 TGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTTGTTGCTTGCTTGGT 19351

Qy        600 AGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 659
                 || || | || || | |   || || | || || | |   || || |  | || |
Db      19352 TCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTCGTTGAT 19411

Qy        660 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTG 719
               |   || || | || |||| |   || || ||| |   ||| ||| ||| |||||  |
Db      19412 TGCTTGGTTGGTTGGTTGCTTGGTTGGTTGGCTGGTTCATTGGTTGGTTGGTTGCTTGGT 19471

Qy        720 GGGCCGCTTGCCTGCT 735
                |   | |||  || |
Db      19472 TGTTTGGTTGGTTGGT 19487


RESULT 24
AP000187
LOCUS       AP000187               100000 bp    DNA     linear   PRI 08-JAN-2000
DEFINITION  Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region,
            clone Q78C10-f32E9, segment 14/21, complete sequence.
ACCESSION   AP000187
VERSION     AP000187.1  GI:4827086
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE     Homo sapiens 2,083,744bp genomic DNA of 21q22.1 (REGION:
            D21S226-AML CLONE RANGE: Q78C10-f32E9)
  JOURNAL   Published Only in DataBase (1999)
REFERENCE   2  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     E. coli transposon insertion:The present data does not contain E.
            coli transposon sequences which integrated in the
            original/previous sequences.  We determined the boundary between
            the insertion and genomic sequences experimentally, removed the
            insertion sequences, reconstituted the present data. The sequencing
            project is supported by Japan Science Technology Corporation (JST)
            and The Institute of Physical and Chemical Research (RIKEN).
FEATURES             Location/Qualifiers
     source          1. .100000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
BASE COUNT    27992 a  22056 c  21621 g  28331 t
ORIGIN

  Query Match             19.3%;  Score 192.8;  DB 9;  Length 100000;
  Best Local Similarity   59.2%;  Pred. No. 4.7e-35;
  Matches  329;  Conservative    0;  Mismatches  227;  Indels    0;  Gaps    0;

Qy        180 TGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGC 239
              || ||| |   | |  |||||| |  |  |   |   |  || | |  || | |   ||
Db      18932 TGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTTGGCTGGTTGCTTGGT 18991

Qy        240 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 299
              || | ||| |  |||||| ||| | | | |  |||||||||||| ||| || ||||||||
Db      18992 TGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 19051

Qy        300 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 359
              |||| ||| || ||  || ||| | | | |  |||||| ||| | | | || ||||||||
Db      19052 TGCTTGCTTGCTTGGCTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGCTTGCTTGCT 19111

Qy        360 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              || | ||| || |  ||| ||||| | | || |||||  ||| | ||| |  ||||||||
Db      19112 TGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCT 19171

Qy        420 GAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGAT 479
                |  |   |   || || ||||||| |  |  |   |   |  ||  || | | ||| |
Db      19172 TCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGCT 19231

Qy        480 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 539
               |   || |||| || |||| |   || || | || || | |'  || |||| || || |
Db      19232 TGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGTTTGGTTGCTTGGTTGGT 19291

Qy        540 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 599
               |   || || | |  || | |   || |  | |  |||| |    | |||| |  || |
Db      19292 TGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTTGTTGCTTGCTTGGT 19351

Qy        600 AGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 659
                 || || | || || | |   || || | || || | |   || || |  | || |
Db      19352 TCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTCGTTGAT 19411

Qy        660 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTG 719
               |   || || | || |||| |   || || ||| |   ||| ||| ||| |||||  |
Db      19412 TGCTTGGTTGGTTGGTTGCTTGGTTGGTTGGCTGGTTCATTGGTTGGTTGGTTGCTTGGT 19471

Qy        720 GGGCCGCTTGCCTGCT 735
                |   | |||  || |
```

```
Db      19472 TGTTTGGTTGGTTGGT 19487


RESULT 25
AP001716
LOCUS       AP001716               340000 bp    DNA     linear   PRI 21-MAY-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 21q, section 60/105.
ACCESSION   AP001716 AL163261 BA000005
VERSION     AP001716.1  GI:7768717
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
            Asakawa,S., Shintani,A., Sasaki,K., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
            Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
            Hennig,S., Riesselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K.,
            Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
            Yaspo,M.L.
  TITLE     The DNA sequence of human chromosome 21
  JOURNAL   Nature 405 (6784), 311-319 (2000)
  MEDLINE   20289799
   PUBMED   10830953
REFERENCE   2  (bases 1 to 340000)
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
            Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
            Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
            Hennig,S., Riesselmann,L., Dagand,E., Wehrmaeyer,S., Borzym,K.,
            Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
            Yaspo,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing
            Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
            Group * Institute of Molecular Biotechnology, Genome Analysis *
            Keio University School of Medicine, Dept. of Molecular Biology *
            GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
            Genetics (addresses see below)
COMMENT     On May 30, 2000 this sequence version replaced gi:7717330.
            The chromosome 21 mapping and sequencing consortium consisting of
            * RIKEN Genomic Sciences Center, Human Genome Research Group, *
            Sagamihara 228-8555, Japan,
            * e.mail: hattori@gsc.riken.go.jp
            * URL: http://hgp.gsc.riken.go.jp/
            and
            * Institute of Molecular Biotechnology, Genome Analysis, *
            Beutenbergstrasse 11, D-07745 Jena, Germany,
            * e.mail: gscj-submit@genome.imb-jena.de
            * URL: http://genome.imb-jena.de/
            and
            * Keio University School of Medicine, Molecular Biology, * Tokyo
            160-8582, Japan,
            * e.mail: nshimizu@dmb-med.keio.ac.jp
            * URL: http://www.dmb.med.keio.ac.jp/
            and
            * GBF, Dept. of Genome Analysis,
            * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
            info.genome@gbf.de
            * URL: http://genome.gbf.de/
            and
            * Max-Planck Institute for Molecular Genetics,
            * Ihnestrasse 73, D-14195 Berlin, Germany,
            * e.mail: info-chr21@molgen.mpg.de
            * URL: http://chr21.rz-berlin.mpg.de/
            AL163261: Submitted (10-Apr-2000).
FEATURES             Location/Qualifiers
     source          1. .340000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
     source          <1. .16929
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="pQ57C2, 5' partial"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000288"
     source          16360. .103735
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="pS701"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000289"
     source          103356. .120643
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="pQ14E2"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000290"
     source          118735. .145947
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="pQ64C8"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000291"
     source          144738. .182946
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="pQ95D4"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000292"
     source          182189. .189406
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="pQ20D3"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000293"
     source          189053. .205485
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
```

```
                     /clone="pD5C10"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000294"
     source          204813. .260943
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="pS318"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000295"
     source          260219. .278406
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="pD71A4"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000296"
     source          277776. .300712
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="pD5C11"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000297"
     source          300164. .325723
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="PQ38G8"
                     /clone_lib="RP1 PAC library"
                     /note="Accession No. AP000298"
     source          322992. .>340000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="PQ102G11, 3' partial"
                     /clone_lib="RP1 PAC library"
                     /note="Accession No. AP000299"
     repeat_region   complement(722. .1108)
                     /note="MLT1B"
                     /rpt_family="LTR/MaLR"
                     /rpt_type=DISPERSED
     repeat_region   complement(1428. .1508)
                     /note="MIR"
                     /rpt_family="SINE/MIR"
                     /rpt_type=DISPERSED
     repeat_region   1794. .1944
                     /note="MIR"
                     /rpt_family="SINE/MIR"

  Query Match             19.3%;  Score 192.8;  DB 9;  Length 340000;
  Best Local Similarity   59.2%;  Pred. No. 4.3e-35;
  Matches  329;  Conservative    0;  Mismatches  227;  Indels    0;  Gaps    0;

Qy        180 TGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGC 239
              || ||| |   | |  |||||| |  |  |   |  || | |  || | |   ||
Db     139106 TGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTTGGCTGGTTGCTTGGT 139165

Qy        240 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 299
              || | ||| |  |||||| ||| | | | |  ||||||||||||| ||| || ||||||||
Db     139166 TGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 139225

Qy        300 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 359
              |||| ||| || ||  || ||| | | | |  |||||| ||| | | | || ||||||||
Db     139226 TGCTTGCTTGCTTGGCTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGCTTGCTTGCT 139285

Qy        360 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              || | ||| || |  ||| ||||| | | || |||||  ||| | ||| |  ||||||||
Db     139286 TGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCT 139345

Qy        420 GAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGAT 479
                |   |   |   || || ||||||| |  |   |   |   |  ||  || | | ||| |
Db     139346 TCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGCT 139405

Qy        480 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 539
                |   || |||| || |||| |   || || | || || | ||  || |||| || || |
Db     139406 TGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGTTTGGTTGCTTGGTTGGT 139465

Qy        540 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 599
                |   || || | |  || | |   || |  | |  |||| |    | |||| |  || |
Db     139466 TGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTTGTTGCTTGCTTGGT 139525

Qy        600 AGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 659
                || || | || || | |   || || | || || | |   || || |  | || |
Db     139526 TCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTCGTTGGAT 139585

Qy        660 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTG 719
                |   || || | || |||| |   || || ||| |   ||| ||| ||| |||||
Db     139586 TGCTTGGTTGGTTGGTTGCTTGGTTGGTTGGCTGGTTCATTGGTTGGTTGGTTGCTTGGT 139645

Qy        720 GGGCCGCTTGCCTGCT 735
                |   | |||  || |
Db     139646 TGTTTGGTTGGTTGGT 139661


RESULT 26
AC094809
LOCUS       AC094809              234939 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-4C12, *** SEQUENCING IN PROGRESS ***,
            2 unordered pieces.
ACCESSION   AC094809
VERSION     AC094809.6  GI:30466564
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 234939)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
```

```
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 234939)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 234939)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 9, 2003 this sequence version replaced gi:24941692.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GBKX
                Center clone name: CH230-4C12
            -------------- Summary Statistics
                Assembly program: Atlas;
                Consensus quality: 216649 bases at least Q40
                Consensus quality: 220061 bases at least Q30
                Consensus quality: 222335 bases at least Q20
                Estimated insert size: 232189; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   231801: contig of 231801 bp in length
            *    231802   231901: gap of unknown length
            *    231902   234939: contig of 3038 bp in length.
FEATURES             Location/Qualifiers
     source          1. .234939
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-4C12"
     misc_feature    1. .1673
                     /note="wgs_contig"
     misc_feature    226281. .228013
                     /note="wgs_contig"
     misc_feature    230341. .231801
                     /note="wgs_contig"
BASE COUNT    71368 a  44035 c  42626 g  65688 t  11222 others
ORIGIN

  Query Match             19.2%;  Score 192.2;  DB 2;  Length 234939;
  Best Local Similarity   57.4%;  Pred. No. 6.1e-35;
  Matches  347;  Conservative    0;  Mismatches  258;  Indels    0;  Gaps    0;

Qy         68 CGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 127
              | ||  | |||||||  |   ||||| |||      | | |   |||||| |||  |
Db     172236 CAGCGCCATGCCTGCCTGTGTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTG 172295

Qy        128 GGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 187
               |||||| |||  |   | | |  ||||||  ||    | ||| |   |||||| |||
Db     172296 TGCCTGCCTGCCTACCTGCCTGCCTGCCAGCCTGTGTGACTGCCTGCCTGCCTGCC 172355

Qy        188 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCT 247
               ||  |||||| |||       | | |   | ||||  || ||   |   | ||||  ||
Db     172356 TGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTGTGCCTGCCTGCCTGCC 172415

Qy        248 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 307
               |||||| ||| |||  ||  || ||| ||| |||  ||  |||||| ||| |||  ||
Db     172416 TGCCTGCCTGCCTGCCTGCCTGCGTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCC 172475

Qy        308 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 367
               |||||| ||| |||  ||  |||||  ||| |||. ||  |||||| ||| |||  ||
Db     172476 TGCCTGCCTGCCTGCCTGCCTGCCTGTGTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCC 172535

Qy        368 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGAC 427
               |||||| ||| |||  ||  |||||| ||| ||   ||  |||||| |||        | |
Db     172536 TGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTGTGCCTGCCTGCCTGCCTGCC 172595

Qy        428 AGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGC 487
               ||   | ||| ||| |||      | |     ||||||||  || ||   ||  | |
Db     172596 AGCCTGTGTGCCTGCCTGCCTGCCTGCCTACCTGCCTGCCTGCCTGCCTGCCTGTGTGCC 172655

Qy        488 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 547
               |||  | || |  |   | ||||  | ||||  |   | ||||  | ||||  |   | |
Db     172656 TGCCTGCCTACCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCC 172715

Qy        548 TGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 607
               ||   | ||||  |   | || |  | ||||  |     |||   | |||   |   | |
Db     172716 TGTGTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCTGTGTGCCTGTGTGCCTGCC 172775

Qy        608 TCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 667
               | |  | ||||  |   | ||||  | ||||  |   | ||||||   |||||   |
Db     172776 TGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCTGCCATGCTACCTGCCA 172835

Qy        668 TGCTG 672
               || ||
Db     172836 TGATG 172840


RESULT 27
AC094484/c
LOCUS       AC094484              235855 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-4B15, WORKING DRAFT SEQUENCE.
```

```
ACCESSION   AC094484
VERSION     AC094484.6  GI:30466820
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 235855)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 235855)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 235855)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 9, 2003 this sequence version replaced gi:24818868.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GASZ
                Center clone name: CH230-4B15
            -------------- Summary Statistics
                Assembly program: Atlas;
                Consensus quality: 224879 bases at least Q40
                Consensus quality: 227006 bases at least Q30
                Consensus quality: 228555 bases at least Q20
                Estimated insert size: 237610; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   235855: contig of 235855 bp in length.
FEATURES             Location/Qualifiers
     source          1. .235855
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-4B15"
     misc_feature    2767. .3585
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH303847"
     misc_feature    complement(231726. .232568)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH303848"
BASE COUNT    61996 a  50589 c  50777 g  66609 t   5884 others
ORIGIN

  Query Match             19.2%;  Score 191.6;  DB 2;  Length 235855;
  Best Local Similarity   56.4%;  Pred. No. 8.5e-35;
  Matches  398;  Conservative    0;  Mismatches  304;  Indels    4;  Gaps    2;

Qy          44 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 103
               ||| ||| |||  | |     ||  | ||||||||||||  ||  ||||||  |||
Db      186287 TGCCTGCCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGTC 186228

Qy         104 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG---GACAGTAGGCCTGCTT 160
                | | |   ||||   |||  ||  | |||| ||||||  |   | | |  |||||  |
Db      186227 TGCCTGCCTGCCTTTCTGCCTGCCTGTCTGCCTGCTGACTGCCTGCCTTTCTGCCTGTCT 186168

Qy         161 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 220
               |       | | |  |||||| ||   ||  | |||  |||      | | |   | ||
```

```
Db     186167 GTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGGCTGCCTGCCTGCCTGCCTGTCT 186108

Qy        221 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 280
              |   || |||   |  | |||||  |||||  |||| || |||  | |||  ||| |||
Db     186107 GTCTGCCTGCCTTTCTGCCTGCTGACTGGCTGGCTTTCTGGCTGTCCGGCTGGCTGGCTG 186048

Qy        281 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 340
              ||| ||||||   || ||| ||| ||||||  ||| |||  ||  |||||  |||  ||
Db     186047 GCTGGCTGGCTGTCTGGCTGGCTGGCTGGCTGGCTGGCTGCCTGCCTGGCTGGCTGACTG 185988

Qy        341 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
              ||  ||| |  ||| ||| ||| ||| |||||  ||| ||   ||  ||||| ||| |
Db     185987 CCTGCCTGCCTGGCTGGCTGGCTGGCT-GCCTGGCTGCCTGGCTGCCTGCCTGGCTGCCT 185929

Qy        401 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTAC 460
              ||  ||  ||||  |||  |  | | |  | |||| ||| |||      | | ||
Db     185928 GCCTGCCTGCCTTGCTGCCTGGCTGCCTGCCTGACTGCCTGCCTGCCTGCCTGTCTGTCT 185869

Qy        461 GCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 520
              |||||||||   || ||   |     ||||  | |||   |   | ||||  | ||||
Db     185868 GCCTGCCTGCCTGCCTGCCTGCCTTTCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGCCT 185809

Qy        521 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTA 580
              ||  | ||||  | ||||  ||  | ||||  | ||||   |  | ||    | ||||
Db     185808 GTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTTTCTGCCTGTCTGTCTGCCT 185749

Qy        581 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 640
              |  ||||  | ||||   |  | ||  |  | ||||  |    | ||||  | ||||
Db     185748 ACCTGCCTGCCTGTCTGCCTTTCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGCCT 185689

Qy        641 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTT 700
              |   | ||||  | ||||  |   | ||   | || |  |   | |||||||   || |
Db     185688 GCCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTGCCTACCTGCCTGCCTGCCTGTCTGCCT 185629

Qy        701 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGC 746
              || ||  ||| |||   |  | | |  || ||||      | ||||
Db     185628 GCCTGACTGCCTGCCTTTCTGCCTGTCTGTCTGCCTGCCTGCCTGC 185583


RESULT 28
AL935202/c
LOCUS       AL935202              189096 bp    DNA     linear   VRT 01-MAY-2003
DEFINITION  Zebrafish DNA sequence from clone CH211-197E11, complete sequence.
ACCESSION   AL935202
VERSION     AL935202.7  GI:30314908
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 189096)
  AUTHORS   Leongamornlert,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 1, 2003 this sequence version replaced gi:28446340.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Repeat names
            beginning 'Dr' were identified by the Recon repeat discovery system
            (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr'
            were identified by Rick Waterman (Stephen Johnson lab, WashU).  For
            further information see
            http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml
            CH211-197E11 is from a CHORI-211 BAC library
            VECTOR: pTARBAC2.1.
FEATURES             Location/Qualifiers
     source          1. .189096
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-197E11"
                     /clone_lib="CHORI-211"
BASE COUNT    58203 a  34096 c  34460 g  62337 t
ORIGIN

  Query Match           19.1%;  Score 191.4;  DB 5;  Length 189096;
  Best Local Similarity   54.9%;  Pred. No. 9.6e-35;
  Matches  378;  Conservative    0;  Mismatches  311;  Indels    0;  Gaps    0;

Qy         69 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 128
              || || |||| |||| | | |  || ||| |  |  |    |        | ||||| |||
Db      99704 GGTTGACTGCTTGCTTGGTTGTTTGGTTGATTGGTTGGTTGGTTAGTCGGTTGCTTGCTT 99645

Qy        129 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 188
              || |||| | |      |    |   |  || |||||     |   |   |  || |||||
Db      99644 GCTTGCTAGTTTGTTTGTTTGCTTGGTTGATTGCTTGCTTGGTTGGTTGTTTGGTTGCTT 99585

Qy        189 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 248
              | | || || ||||| ||  ||  |   || || | ||||| | |   || ||     |
Db      99584 GTTTGCTTGGTTGCTCAGTTGATTGCTTGGTTGATAGCTTGGTTGGTTGGTTGGGTAGTT 99525

Qy        249 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 308
              || |||||| ||||| | | |  || ||| ||| | | | |  |||||| ||  | | |
Db      99524 GCTTGCTTGTTTGCTTGGTTGGTTGGTTGGTTGTTTGGTTGATTGCTTGGTTCGTTGGTT 99465

Qy        309 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 368
              |  || ||| ||| | | | |  | |||| ||| | | | |  ||||||  ||  |   |
Db      99464 GATTGGTTGGTTGTTTGGTTGATTTCTTGGTTGGTTGGTTGATTGCTTGGTTAGTTTATT 99405

Qy        369 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 428
              |  || ||  || |||||| || |||||| ||||| | | |  || ||| | ||  |
Db      99404 GGTTGGTTAGTTTCTCGCTTGCTTGCTTGTTTGCTTGGTTGGTTGGTTGGTTAGTTGCTT 99345

Qy        429 GTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCT 488
              |   | || |||||||  | ||  |   |   || ||  || | | ||||| |   || |
Db      99344 GCTTGTTTGTTTGCTTGTTTAGTTGGTTGGTTGCATGTTTGCTTGGTTGATTGCTTGGTT 99285

Qy        489 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 548
                | || || | |   || || |  | ||   |   ||  | | |  || | |  ||  |
Db      99284 ATTTGGTTGGTTGGCTGGTTGGTTAGTTGGATGGTTGGTCGGTTGCTTGGTTGATGGCTT 99225

Qy        549 GCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 608
              ||| |  || | |   |  | || || | |  | ||||| ||  | | |   | |  |
Db      99224 GCTTGCGTGTTTGCTTGCTTTGTTGGTTGGTTGGTTGACTGCTTGGTCGGTTGGTTGATT 99165

Qy        609 CCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 668
              || || || |      |  || | || || | ||  || || | |  || | |   || |
Db      99164 ACTTGGTTGGTTAGTTGATTGGTTGGTTGGTTGTTTGGTTGATTGCTTGGTTGGTTGGTT 99105

Qy        669 GCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTT 728
              | | |  |||| |   || || ||| | | ||| | | ||||||| | |   ||  ||||
```

```
Db      99104 GGTTGCTTGCTTGGTTGGTTGGCTGGTTGTTTGGTCGATTGCTTGGTTGGTTGGTTGCTT 99045

Qy        729 GCCTGCTACTAGGGCTGCTGTGCAGCTGG 757
              |  || |     || || |  |  | | |
Db      99044 GGTTGGTTGGCTGGTTGTTTGGTTGATTG 99016


RESULT 29
AC112928
LOCUS       AC112928               223156 bp    DNA     linear   HTG 18-MAR-2003
DEFINITION  Mus musculus clone RP24-275P22, WORKING DRAFT SEQUENCE, 8 unordered
            pieces.
ACCESSION   AC112928
VERSION     AC112928.4  GI:29029390
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 223156)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-275P22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 223156)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 223156)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 18, 2003 this sequence version replaced gi:28209605.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
               Center project name: L18718
               Center clone name: 275_P_22
            -------------- Summary Statistics
               Sequencing vector: Plasmid; n/a; 100% of reads
               Chemistry: Dye-terminator.Big Dye; 100% of reads
               Assembly program: Phrap; version 0.960731
               Consensus quality: 221447 bases at least Q40
               Consensus quality: 222092 bases at least Q30
               Consensus quality: 222332 bases at least Q20
               Insert size: 194000; agarose-fp
               Insert size: 222456; sum-of-contigs
               Quality coverage: 11.2 in Q20 bases; agarose-fp
               Quality coverage: 9.8 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     5792: contig of 5792 bp in length
            *    5793     5892: gap of 100 bp
            *    5893     6990: contig of 1098 bp in length
            *    6991     7090: gap of 100 bp
            *    7091    12953: contig of 5863 bp in length
            *   12954    13053: gap of 100 bp
            *   13054    34245: contig of 21192 bp in length
            *   34246    34345: gap of 100 bp
            *   34346    60533: contig of 26188 bp in length
            *   60534    60633: gap of 100 bp
            *   60634   100582: contig of 39949 bp in length
            *  100583   100682: gap of 100 bp
            *  100683   192104: contig of 91422 bp in length
            *  192105   192204: gap of 100 bp
            *  192205   223156: contig of 30952 bp in length.
FEATURES             Location/Qualifiers
     source          1. .223156
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-275P22"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .5792
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    5893. .6990
                     /note="assembly_fragment"
     misc_feature    7091. .12953
                     /note="assembly_fragment"
     misc_feature    13054. .34245
                     /note="assembly_fragment"
     misc_feature    34346. .60533
                     /note="assembly_fragment"
     misc_feature    60634. .100582
                     /note="assembly_fragment"
     misc_feature    100683. .192104
                     /note="assembly_fragment"
     misc_feature    192205. .223156
                     /note="assembly_fragment
                     clone_end:T7
```

```
                    vector_side:right"
BASE COUNT    67283 a  45092 c  44955 g  65105 t    721 others
ORIGIN

  Query Match            19.1%;  Score 190.6;  DB 2;  Length 223156;
  Best Local Similarity  71.6%;  Pred. No. 1.5e-34;
  Matches  250;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy         72 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 131
              ||| ||| |||| ||| || |||||||| |   |  || |||||||| ||| ||
Db     150073 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCGTGCTTGCTTGCTTGCG 150132

Qy        132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191
              ||||||||     |   |   || ||||||||     |   |   || |||||||| |||
Db     150133 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCGTGCTTGCTTGCT 150192

Qy        192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251
               || |||||||      |   |   | |||| ||||||| |   |  |||| ||  ||
Db     150193 TGCTTGCTTGCGTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCGTGCG 150252

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
              |||||||||||| ||| || |||||||||||| ||| || |||||||||||| ||  ||
Db     150253 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCGTGCT 150312

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
              |||||||||||| ||  || |||||||||||| |||  | |||||||||||| ||| ||
Db     150313 TGCTTGCTTGCTTGCGTGCTTGCTTGCTTGCTTGCTTACGTGCTTGCTTGCTTGCTT 150372

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 420
              |||||||||||| ||| || |||||||||||| ||| || |||||||||
Db     150373 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCGTGCTTGCTG 150421


RESULT 30
AL844846
LOCUS       AL844846              209013 bp    DNA     linear   ROD 05-MAR-2003
DEFINITION  Mouse DNA sequence from clone RP23-358I19 on chromosome 2, complete
            sequence.
ACCESSION   AL844846
VERSION     AL844846.11  GI:27902000
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 209013)
  AUTHORS   Bates,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 25, 2003 this sequence version replaced gi:25955774.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to confirm this sequence.  Sequence data
            from the whole genome shotgun alone has only been used where it has
            a phred quality of at least 30.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-358I19 is
            from the RPCI-23 Mouse BAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .209013
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-358I19"
                     /clone_lib="RPCI-23"
BASE COUNT    59522 a  49671 c  46169 g  53651 t
ORIGIN

  Query Match            18.9%;  Score 189.4;  DB 10;  Length 209013;
  Best Local Similarity  63.5%;  Pred. No. 2.8e-34;
  Matches  289;  Conservative    0;  Mismatches  166;  Indels    0;  Gaps    0;

Qy         52 TGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTA 111
              ||| |   |  ||  |  ||| |||||||| ||| || ||||||||     |   |
Db     152496 TGTTTGCTTGCTTGTTTGCTTGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCT 152555

Qy        112 GGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGAC 171
               || |||||||| ||| |  ||||||||     |   |   || ||||||||     |
Db     152556 TGCTTGCTTGCTTGCTTGTTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 152615

Qy        172 AGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCT 231
               ||  || |||||||| | | || || ||| |     |   ||  |  |||| |||||||
Db     152616 TGTTTGCTTGCTTGCTTGTTTGCTTGTTTGTTTGCTTGTTTGTTTGCTTGCTTGCTTGCT 152675

Qy        232 AGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 291
               ||  |  |||| ||| || || ||||||||| ||| |  ||||||||| | | | |
Db     152676 TGTTTGCTTGCTTGCTTGCTTGTTTGCTTGCTTGCTTGTTTGCTTGCTTGTTTGTTTGTT 152735

Qy        292 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 351
              |||||||||||| | | || |||||||||||| | | || |||||||||||| ||| |
Db     152736 TGCTTGCTTGCTTGTTTGCTTGCTTGCTTGCTTGTTTGCTTGCTTGCTTGCTTGCTTGTT 152795

Qy        352 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
               || |||||||||   | |  || ||||||| | ||| || |||||| ||| | | | ||
Db     152796 TGTTTGCTTGCTTATTTGTTTGTTTGCTTGTTTGCTTGCTTGCTTGTTTGTTTGTTTGCT 152855

Qy        412 TGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGAT 471
              ||||||  |     |   |   | ||||||||||||     |   |   || ||  || |
Db     152856 TGCTTGTTTGCTTGCTTGCTTGTTTGCTTGCTTGCTTGTTTGCTTGCTTGTTTGTT 152915

Qy        472 GGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGG 506
              ||||| | ||  |  || | | ||| || | | |
Db     152916 TGCTTGCTTGTTTGTTTGTTTGTCTGTTACTGGTG 152950


RESULT 31
AC114591/c
LOCUS       AC114591              208740 bp    DNA     linear   HTG 24-FEB-2003
DEFINITION  Mus musculus clone RP23-478G2, WORKING DRAFT SEQUENCE, 7 ordered
            pieces.
ACCESSION   AC114591
VERSION     AC114591.5  GI:28475804
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 208740)
```

```
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-478G2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 208740)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 208740)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 24, 2003 this sequence version replaced gi:21392539.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L24156
                Center clone name: 478_G_2
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 207158 bases at least Q40
                Consensus quality: 207792 bases at least Q30
                Consensus quality: 208011 bases at least Q20
                Insert size: 210000; agarose-fp
                Insert size: 208140; sum-of-contigs
                Quality coverage: 11.2 in Q20 bases; agarose-fp
                Quality coverage: 11.3 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    11975: contig of 11975 bp in length
            *   11976    12075: gap of 100 bp
            *   12076    14622: contig of 2547 bp in length
            *   14623    14722: gap of 100 bp
            *   14723   125098: contig of 110376 bp in length
            *  125099   125198: gap of 100 bp
            *  125199   139380: contig of 14182 bp in length
            *  139381   139480: gap of 100 bp
            *  139481   158709: contig of 19229 bp in length
            *  158710   158809: gap of 100 bp
            *  158810   195661: contig of 36852 bp in length
            *  195662   195761: gap of 100 bp
            *  195762   208740: contig of 12979 bp in length.
FEATURES             Location/Qualifiers
     source          1. .208740
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-478G2"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .11975
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    12076. .14622
                     /note="assembly_fragment"
     misc_feature    14723. .125098
                     /note="assembly_fragment"
     misc_feature    125199. .139380
                     /note="assembly_fragment"
     misc_feature    139481. .158709
                     /note="assembly_fragment"
     misc_feature    158810. .195661
                     /note="assembly_fragment"
     misc_feature    195762. .208740
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    59319 a  44577 c  46204 g  58040 t    600 others
ORIGIN

  Query Match             18.9%;  Score 189.2;  DB 2;  Length 208740;
  Best Local Similarity   56.4%;  Pred. No. 3.1e-34;
  Matches  353;  Conservative    0;  Mismatches  273;  Indels    0;  Gaps    0;

Qy         80 TGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCT 139
              |||| | | ||  |  | ||| |  |   |   ||  | ||||| | | ||  | |||||
Db     148302 TGCTGGGTTGCTGGGATACTGGGTTGCTGGGTTGCTGGGTTGCTGGGTTGCTGGGTTGCT 148243

Qy        140 GAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCT 199
              | |  |  ||   ||  |  | ||| | ||   |  |||  | ||||| |    |  | |
Db     148242 GGGTTGCTAGGTTGCTGGGATACTGGGTGGCTGGGTGGCTGGGTTGCTGGGATACTGGGT 148183

Qy        200 TGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCT 259
              ||||| | ||   |  ||   | ||||| | | |  |||  ||| | ||||  | |||||
Db     148182 TGCTGGGTGGCTGGGTGGCTGGGTGGCTGGGTGGCTGGGTGGCTGGATGGCTGGGTTGCT 148123

Qy        260 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 319
               | | |||||  |||| | || || | | ||  | ||||| | | |||||  | || | |
Db     148122 GGGTTGCTGGGTTGCTGGGTTACTGGGTTGCTGGGTTGCTAGGTTGCTGGGATACTGGGT 148063
```

```
Qy        320 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 379
              ||| | | ||  | ||||| | | |||||  | || | ||||| | | ||  | |||||
Db     148062 GGCTGGGTTGCTGGGTTGCTGGTTTGCTGGGTTACTGGGTTGCTGGGTTGCTGGGTTGCT 148003

Qy        380 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 439
              | | |||||  | || | | ||| | | ||  | |||||| |      |   |   | |
Db     148002 AGGTTGCTGGGATACTGGGTGGCTGGGTTGCTGGGTTGCTGGGTTACTGGGTTGCTGGGT 147943

Qy        440 TGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCT 499
              |||| | | |  ||   |   | |||  | | ||| | | |  ||| | ||||  |||
Db     147942 TGCTGGGTTACTGGGTTGCTGGGTTGCTGGGTTGCTAGGTTGCTGGGATACTGGGTGGCT 147883

Qy        500 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 559
              |  ||   | |||   | | |  ||| | ||||| |||| |  ||   | |||   | |
Db     147882 GGGTGGCTGGGTGGCTGGGTTGCTGGGATACTGGGTTGCTGGGTGGCTGGGTGGCTGGGT 147823

Qy        560 AGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCT 619
              |  ||    ||||  ||| |  ||   | |||   | |||  ||| | ||||| | ||
Db     147822 GGCTGGATGGCTGGGTGGCTGGGTGGCTGGGTGGCTGGGTAGCTGGGTTGCTGGGATACT 147763

Qy        620 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 679
              |   |   | |||   | | |  ||| ||||||  ||| |  ||   | |||   | |
Db     147762 GGGTTGCTGGGTGGCTGGGTGGCTGGGTGGCTGGGTGGCTGGGTGGCTGGGTGGCTGGGT 147703

Qy        680 AGTAGGGCTGCCTGCTGGCTTGCTTG 705
              |  ||| |||  | | ||| | | |
Db     147702 GGCTGGGATGCTGGGTTGCTGGTTGG 147677
```

```
RESULT 32
AC097040
LOCUS       AC097040               230500 bp    DNA     linear   HTG 13-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-66A9, WORKING DRAFT SEQUENCE.
ACCESSION   AC097040
VERSION     AC097040.7  GI:24942620
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 230500)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 230500)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 230500)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 13, 2002 this sequence version replaced gi:23664821.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GHYK
                Center clone name: CH230-66A9
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 218247 bases at least Q40
                Consensus quality: 220700 bases at least Q30
                Consensus quality: 222115 bases at least Q20
                Estimated insert size: 223843; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
```

```
     *          1   230500: contig of 230500 bp in length.
FEATURES        Location/Qualifiers
     source     1. .230500
                /organism="Rattus norvegicus"
                /mol_type="genomic DNA"
                /db_xref="taxon:10116"
                /clone="CH230-66A9"
     misc_feature    1. .1087
                /note="wgs_end_extension
                clone_end:T7"
     misc_feature    2577. .3456
                /note="clone_boundary
                clone_end:T7
                site:
                end_sequence:BH264043"
BASE COUNT    57052 a  55956 c  55370 g  54914 t   7208 others
ORIGIN

  Query Match           18.8%;  Score 187.6;  DB 2;  Length 230500;
  Best Local Similarity   53.7%;  Pred. No. 7.4e-34;
  Matches  388;  Conservative    0;  Mismatches  334;  Indels    0;  Gaps    0;

Qy         25 GGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTC 84
Db     101774 GAGGTCAGGGAGGGACTTTCTGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 101833

Qy         85 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 144
Db     101834 GCCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCT 101893

Qy        145 GACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 204
Db     101894 GTCTGCCTGTCTGCCTGCCTGTCTGCCTGTCTCTCTGCCTGCCTGCCTGTCTGCCTGTCT 101953

Qy        205 AGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTC 264
Db     101954 GCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGCCTGTCTGTCTGTCTGCCTGTCT 102013

Qy        265 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 324
Db     102014 GCCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCT 102073

Qy        325 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 384
Db     102074 GCCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGTCTGTCTGTCTGTCTGTCTGCCTGTCT 102133

Qy        385 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 444
Db     102134 GCCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCT 102193

Qy        445 GCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAG 504
Db     102194 GCCTGCCTGCCTGTCTGCCTGTCTGTCTGTCTGTCTGCCTGTCTGCCTGCCTGCCT 102253

Qy        505 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 564
Db     102254 GTCTGCCTGTCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTGCCTGCCT 102313

Qy        565 GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 624
Db     102314 GCCTGCCTGTCTGCCTGTCTGTCTGTCTGTCTGTCTGCCTGTCTGCCTGTCTGCCTGTCT 102373

Qy        625 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 684
Db     102374 GCCTGTCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCCGTCT 102433

Qy        685 GGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCT 744
Db     102434 GCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGCCTGTCTGTCTGCCT 102493

Qy        745 GC 746
Db     102494 GC 102495
```

```
RESULT 33
AC142075/c
LOCUS       AC142075              258132 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-140G17, WORKING DRAFT SEQUENCE, 5
            unordered pieces.
ACCESSION   AC142075
VERSION     AC142075.3  GI:30579162
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 258132)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljević,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 258132)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
```

```
REFERENCE   3  (bases 1 to 258132)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:29469497.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GJTN
                Center clone name: CH230-140G17
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 228398 bases at least Q40
                Consensus quality: 231939 bases at least Q30
                Consensus quality: 234817 bases at least Q20
                Estimated insert size: 242804; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    10188: contig of 10188 bp in length
            *    10189    10288: gap of unknown length
            *    10289   252038: contig of 241750 bp in length
            *   252039   252138: gap of unknown length
            *   252139   253967: contig of 1829 bp in length
            *   253968   254067: gap of unknown length
            *   254068   255789: contig of 1722 bp in length
            *   255790   255889: gap of unknown length
            *   255890   258132: contig of 2243 bp in length.
FEATURES             Location/Qualifiers
     source          1. .258132
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-140G17"
     misc_feature    2429. .5186
                     /note="wgs_contig"
BASE COUNT    59022 a  58568 c  58799 g  60346 t  21397 others
ORIGIN

  Query Match           18.8%;  Score 187.6;  DB 2;  Length 258132;
  Best Local Similarity   53.7%;  Pred. No. 7.3e-34;
  Matches  388;  Conservative    0;  Mismatches  334;  Indels    0;  Gaps    0;

Qy         25 GGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTC 84
              | || |||||   | | |     ||  |||  | |     ||  | ||| ||| |||
Db     149939 GAGGTCAGGGAGGGACTTTCTGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 149880

Qy         85 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 144
              ||  |||||| ||        | | |   |||||| |||  ||  | |||| ||
Db     149879 GCCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCT 149820

Qy        145 GACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 204
              | | |   | |||| |||       | | ||   |||| |||  ||  | |||| ||
Db     149819 GTCTGCCTGTCTGCCTGCCTGTCTGCCTGTCTCTCTGCCTGCCTGCCTGTCTGCCTGTCT 149760

Qy        205 AGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTC 264
                | | |   | ||||  |  |||  ||  | ||||  |   | |||  ||| ||
Db     149759 GCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGCCTGTCTGTCTGTCTGCCTGTCT 149700

Qy        265 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 324
              ||  |||||| ||  |||  |   | |||| ||| |||  |   |||||  ||| ||
Db     149699 GCCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCT 149640

Qy        325 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 384
              ||  | |||| ||  ||   |  ||||||  ||  ||   |  | |||  ||| ||
Db     149639 GCCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGTCTGTCTGTCTGTCTGTCTGCCTGTCT 149580

Qy        385 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 444
              ||  | |||| ||| |||  ||  |||||  |||       | | ||  | |||  ||| |
Db     149579 GCCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCT 149520

Qy        445 GCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAG 504
              ||       | | ||  |||||  |||   |  ||   ||  | |||   | ||||  |
Db     149519 GCCTGCCTGCCTGTCTGCCTGTCTGTCTGTCTGTCTGTCTGCCTGTCTGCCTGCCTGCCT 149460

Qy        505 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 564
              | ||||  | ||||  ||  | ||||  | ||||  ||  | ||||  | ||||  |
Db     149459 GTCTGCCTGTCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTGCCTGCCT 149400

Qy        565 GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 624
              | ||  |  | ||||  ||  | |||   | |||   |   | ||  |  | ||||  ||
Db     149399 GCCTGCCTGTCTGCCTGTCTGTCTGTCTGTCTGTCTGCCTGTCTGCCTGTCTGCCTGTCT 149340

Qy        625 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 684
              | |||   | |||   ||  | ||||  | |||   |   | ||||  | ||||  ||
Db     149339 GCCTGTCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCCGTCT 149280

Qy        685 GGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCT 744
              | |||||||  |  ||| ||  ||| ||   |   | | || ||||||    |  | ||
Db     149279 GCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGCCTGTCTGTCTGCCT 149220

Qy        745 GC 746
              ||
Db     149219 GC 149218


RESULT 34
AC131371/c
LOCUS       AC131371               203152 bp    DNA     linear   HTG 15-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-318L15, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC131371
VERSION     AC131371.4  GI:25007315
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 203152)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
```

```
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 203152)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 203152)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 15, 2002 this sequence version replaced gi:23908406.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ---------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
                Center project name: KCBL
                Center clone name: CH230-318L15
            ---------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 189099 bases at least Q40
                Consensus quality: 191887 bases at least Q30
                Consensus quality: 193446 bases at least Q20
                Estimated insert size: 194163; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   200227: contig of 200227 bp in length
            *    200228   200327: gap of unknown length
            *    200328   203152: contig of 2825 bp in length.
FEATURES             Location/Qualifiers
     source          1. .203152
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-318L15"
     misc_feature    1. .1183
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    2461. .4356
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    5237. .5898
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BZ141702"
     misc_feature    complement(198163. .199006)
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BZ141701"
BASE COUNT    47960 a  45439 c  46827 g  54363 t   8563 others
ORIGIN

  Query Match             18.6%;  Score 186.2;  DB 2;  Length 203152;
  Best Local Similarity   58.8%;  Pred. No. 1.6e-33;
  Matches  376;  Conservative    0;  Mismatches  258;  Indels    5;  Gaps    3;

Qy         85 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 144
              ||  |||||||| ||      |   |  |||||||||| ||  | |||||||||
Db     189398 GCCTGCCTGCTTCCTTGCTTGTTTGCTTGCCTGCTTGCTTGCCTGGCTGCTTGCTTGTTT 189339

Qy        145 GACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 204
              |   |   ||||||||||      ||  |   || |||||||  ||  ||||||||||
Db     189338 GCTTGCCTGCCTGCTTGCTTGTTTGAATGCTTGCTTGCTTGCCTGCCTGCCTGCTTGCTT 189279

Qy        205 AGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTC 264
                ||  |   |  |||| || |||  |   |  || ||||| || | |||||| |||
Db     189278 GTTTGAATGCTTGCTTGCTTGCCTGCCTGCCTGCTTGTTCGCTTGCTTACTTGCTCGCTT 189219

Qy        265 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 324
              ||| || ||||||||||  ||| |  | || ||||||||||    ||||||||||||
Db     189218 GCTCGCTTGCTTGCTTGCATGCTTGATTCCTGGCTTGCTCGCTTAAATGCTTGCTTGCTC 189159

Qy        325 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 384
              |||   .|| |||||||||   | |  ||||||||||| ||  |||||  | | ||||
Db     189158 GCTTAAATGATTGCTTGCTTATTTGAATGCTTGCTTGCTTGCCTGCCTGACTACCTGCTT 189099

Qy        385 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 444
              ||| ||    |||||||||| || || ||||| ||     |   |      |||||||||
```

```
Db     189098 GCTTGC---TTTGCTTGCTCACTTGCTTGCTTCCTTGCTTGCTTGCTTAAATGCTTGCTT 189042

Qy        445 GCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAG 504
              ||| |   |   |   |  ||  || | ||||| | |   | ||||  | ||||  |
Db     189041 GCTCACTTGCTTGCTTGTTTGAATGCTTGCTTGCTCGCTTGCCTGCCTGCCTGCCTGCCT 188982

Qy        505 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT-AGTA 563
              | ||||| |  ||||     |  |||| |  |||| |      |||| |  |||| | |
Db     188981 GCCTGCTTGCTTGCTCACTTGCTTGCTTGAATGCTTGCTTCCTTGCTTGCTTGCTCACTT 188922

Qy        564 GGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTA 623
              |    | ||  ||||  |   |  ||||       |   |  | || |  |||| |
Db     188921 GCTTGCTTGCTTTGCTTGCTTGCTTGCTCACTCAGTTCCTTGCTTGCTTGAATGCTTGCT 188862

Qy        624 GGGCTGCTGGGCTG-CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGT 682
               |  ||||  |||  || |   |    || |  || |     |  |||| |  ||||
Db     188861 TGCTTGCTTTGCTCACTTGCTTGCTCACTTGATTGTTCCCTTGCTTGCTTGTTTGCTCAC 188802

Qy        683 AGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGG 721
                |  ||| ||||  ||  ||||||  |||||| | |||
Db     188801 CTGCTTGCTTGCTCACTCACTTGCTCACTTGCTTGGGGG 188763
```

RESULT 35
AC095180/c
LOCUS AC095180 270245 bp DNA linear HTG 13-NOV-2002
DEFINITION Rattus norvegicus clone CH230-9C14, WORKING DRAFT SEQUENCE, 3 unordered pieces.
ACCESSION AC095180
VERSION AC095180.5 GI:24940977
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 270245)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 270245)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 270245)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 13, 2002 this sequence version replaced gi:22772773.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
-------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
Center project name: GDAH
Center clone name: CH230-9C14
-------------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 230636 bases at least Q40
Consensus quality: 232982 bases at least Q30
Consensus quality: 234595 bases at least Q20
Estimated insert size: 238600; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 74405: contig of 74405 bp in length
* 74406 74505: gap of unknown length
* 74506 196450: contig of 121945 bp in length
* 196451 196550: gap of unknown length
* 196551 270245: contig of 73695 bp in length.
FEATURES Location/Qualifiers
source 1. .270245

```
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-9C14"
     misc_feature    194462. .196450
                     /note="wgs_contig"
BASE COUNT    60901 a  56122 c  55375 g  63441 t  34406 others
ORIGIN

  Query Match           18.6%;  Score 186.2;  DB 2;  Length 270245;
  Best Local Similarity   58.8%;  Pred. No. 1.6e-33;
  Matches  376;  Conservative    0;  Mismatches  258;  Indels    5;  Gaps    3;

Qy         85 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 144
              ||  |||||||| ||      |   |   |||||||||||| ||  | |||||||||
Db     219496 GCCTGCCTGCTTCCTTGCTTGTTTGCTTGCCTGCTTGCTTGCCTGGCTGCTTGCTTGTTT 219437

Qy        145 GACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 204
              |   |   ||||||||||||      ||  |   || |||||||  ||  ||||||||||||
Db     219436 GCTTGCCTGCCTGCTTGCTTGTTTGAATGCTTGCTTGCTTGCCTGCCTGCCTGCTTGCTT 219377

Qy        205 AGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTC 264
                 ||  |   |  |||| || |||  |    |  || ||||| || | |||||| |||
Db     219376 GTTTGAATGCTTGCTTGCTTGCCTGCCTGCCTGCTTGTTCGCTTGCTTACTTGCTCGCTT 219317

Qy        265 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 324
              ||| || ||||||||||  ||| |  | || ||||||||||    ||||||||||||
Db     219316 GCTCGCTTGCTTGCTTGCATGCTTGATTCCTGGCTTGCTCGCTTAAATGCTTGCTTGCTC 219257

Qy        325 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 384
              |||    || |||||||||   | |  ||||||||||| ||  |||||  | | ||||
Db     219256 GCTTAAATGATTGCTTGCTTATTTGAATGCTTGCTTGCTTGCCTGCCTGACTACCTGCTT 219197

Qy        385 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 444
              ||| ||    ||||||||| || || ||||| ||     |   |      |||||||||
Db     219196 GCTTGC---TTTGCTTGCTCACTTGCTTGCTTCCTTGCTTGCTTGCTTAAATGCTTGCTT 219140

Qy        445 GCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAG 504
              ||| |   |   |   |  ||  || | ||||| | |   | ||||  | ||||  |
Db     219139 GCTCACTTGCTTGCTTGTTTGAATGCTTGCTTGCTCGCTTGCCTGCCTGCCTGCCTGCCT 219080

Qy        505 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT-AGTA 563
              | ||||| |  |||      |  |||| |  |||| |      ||||  |  |||| | |
Db     219079 GCCTGCTTGCTTGCTCACTTGCTTGCTTGAATGCTTGCTTCCTTGCTTGCTTGCTCACTT 219020

Qy        564 GGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTA 623
              |    | ||   |||| |   |  ||||       |     |  | || |  |||| |
Db     219019 GCTTGCTTGCTTTGCTTGCTTGCTTGCTCACTCAGTTCCTTGCTTGCTTGAATGCTTGCT 218960

Qy        624 GGGCTGCTGGGCTG-CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGT 682
               |  ||||  |||  || |   |    || |  || |     |  |||| |  ||||
Db     218959 TGCTTGCTTTGCTCACTTGCTTGCTCACTTGATTGTTCCCTTGCTTGCTTGTTTGCTCAC 218900

Qy        683 AGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGG 721
                |  ||| ||||  ||  ||||||  |||||| | |||
Db     218899 CTGCTTGCTTGCTCACTCACTTGCTCACTTGCTTGGGGG 218861


RESULT 36
U93872/c
LOCUS       U93872                133661 bp    DNA     linear   VRL 09-JUL-2001
DEFINITION  Kaposi's sarcoma-associated herpesvirus glycoprotein M, DNA
            replication protein, glycoprotein, DNA replication protein, FLICE
            inhibitory protein and v-cyclin genes, complete cds, and tegument
            protein gene, partial cds.
ACCESSION   U93872
VERSION     U93872.2  GI:14627174
KEYWORDS    .
SOURCE      Human herpesvirus 8 (Kaposi's sarcoma-associated herpesvirus -
            Human herpesvirus 8)
  ORGANISM  Human herpesvirus 8
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Rhadinovirus.
REFERENCE   1  (bases 29032 to 30108; 117733 to 118431)
  AUTHORS   Zhong,W., Wang,H., Herndier,B. and Ganem,D.
  TITLE     Restricted expression of Kaposi sarcoma-associated herpesvirus
            (human herpesvirus 8) genes in Kaposi sarcoma
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 93 (13), 6641-6646 (1996)
  MEDLINE   96270595
   PUBMED   8692871
REFERENCE   2  (bases 17242 to 17856)
  AUTHORS   Neipel,F., Albrecht,J.C., Ensser,A., Huang,Y.Q., Li,J.J.,
            Friedman-Kien,A.E. and Fleckenstein,B.
  TITLE     Human herpesvirus 8 encodes a homolog of interleukin-6
  JOURNAL   J. Virol. 71 (1), 839-842 (1997)
  MEDLINE   97138401
   PUBMED   8985427
REFERENCE   3  (bases 123309 to 124082)
  AUTHORS   Li,M., Lee,H., Yoon,D.W., Albrecht,J.C., Fleckenstein,B., Neipel,F.
            and Jung,J.U.
  TITLE     Kaposi's sarcoma-associated herpesvirus encodes a functional cyclin
  JOURNAL   J. Virol. 71 (3), 1984-1991 (1997)
  MEDLINE   97184528
   PUBMED   9032330
REFERENCE   4  (bases 122660 to 123226)
  AUTHORS   Thome,M., Schneider,P., Hofmann,K., Fickenscher,H., Meinl,E.,
            Neipel,F., Mattmann,C., Burns,K., Bodmer,J.L., Schroter,M.,
            Scaffidi,C., Krammer,P.H., Peter,M.E. and Tschopp,J.
  TITLE     Viral FLICE-inhibitory proteins (FLIPs) prevent apoptosis induced
            by death receptors
  JOURNAL   Nature 386 (6624), 517-521 (1997)
  MEDLINE   97242415
   PUBMED   9087414
REFERENCE   5  (bases 1 to 133661)
  AUTHORS   Neipel,F., Albrecht,J.C. and Fleckenstein,B.
  TITLE     Cell-homologous genes in the Kaposi's sarcoma-associated
            rhadinovirus human herpesvirus 8: determinants of its
            pathogenicity?
  JOURNAL   J. Virol. 71 (6), 4187-4192 (1997)
  MEDLINE   97296220
   PUBMED   9151804
REFERENCE   6  (bases 1 to 133661)
  AUTHORS   Neipel,F., Albrecht,J.-C., Ensser,A., Huang,Y.-Q., Li,J.J.,
            Friedman-Kien,A.E. and Fleckenstein,B.
  TITLE     The genome of human herpesvirus 8 cloned from Kaposi's sarcoma
  JOURNAL   Unpublished
REFERENCE   7  (bases 1 to 133661)
  AUTHORS   Neipel,F., Albrecht,J.-C., Ensser,A., Huang,Y.-Q., Li,J.J.,
            Friedman-Kien,A.E. and Fleckenstein,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1997) Virology, University of Erlangen,
            Schlossgarten 4, Institut for Klinische und Molekulare, Erlangen
            91054, Germany
REFERENCE   8  (bases 1 to 133661)
  AUTHORS   Neipel,F., Albrecht,J.-C., Ensser,A., Huang,Y.-Q., Li,J.J.,
            Friedman-Kien,A.E. and Fleckenstein,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Virology, University of Erlangen,
            Schlossgarten 4, Institut for Klinische und Molekulare, Erlangen
            91054, Germany
  REMARK    Sequence update by submitter
COMMENT     On Jul 9, 2001 this sequence version replaced gi:2246466.
FEATURES             Location/Qualifiers
     source          1. .133661
                     /organism="Human herpesvirus 8"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:37296"
                     /note="sequence derived from overlapping Lambda phage
                     clones isolated from two libraries of Sau3A partially
                     digested DNA from Kaposi's sarcoma biopsies"
     CDS             105. .959
                     /note="ORF K1"
                     /codon_start=1
                     /protein_id="AAB62619.1"
                     /db_xref="GI:2246494"
                     /translation="MFLYVVCSLAVCFRGLLSLSLQSSPNLCPGVISTPYTLTCPSNT
```

```
                SLPTSWYCNDTRLLRLTQQTFTVDTLICNFSCVGQQSGHRHSLWITWYPQPVLQTLCGQ
                PSNTVTCGQHVTLYCSTSGNNVTVWHLPNGRNETVSQTKYYNFTLMNQTEGCYACSNG
                LSSRLSNRLCFSARCANLTPETHTVSVSSTTGFRTFSTNRRGTTLDVLVMKEAKSTNL
                HIQVHFLVFMTLVALIGTMCGILGTIIFAHCQKQRDSNKTVPQQLQDYYSLHDLCTED
                YTQPVDWY"
CDS             1127. .2779
                /note="ORF 04, complement control protein homolog"
                /codon_start=1
                /protein_id="AAB62602.1"
                /db_xref="GI:2246477"
                /translation="MAFLRQTLWTLWTFTMVIGQDNEKCSQKTLIGYRLKMSRDGDIA
                VGETVELRCRSGYTTYARNITATCLQGGTWSEPTATCNKKSCPNPGEIQNGKVIFHGG
                QDALKYGANISYVCNEGYFLVGREYVRYCMIGASGQMAWSSSPPFCEKEKCHRPKIEN
                GDFKPDKDYYEYNDAVHFECNEGYTLVGPHSIACAVNNTWTSNMPTCELAGCKFPSVT
                HGYPIQGFSLTYKHKQSVTFACNDGFVLRGSPTITCNVTEWDPPLPKCVLEDIDDPNN
                SNPGRLHPTPNEKPNGNVFQRSNYTEPPTKPEDTHTAATCDTNCEQPPKILPTSEGFN
                ETTTSNTITKQLEDEKTTSQPNTHITSALTSMKAKGNFTNKTNNSTDLHIASTPTSQD
                DATPSIPSVQTPNYNTNAPTRTLTSLHIEEGPSNSTTSEKATASTLSHNSHKNDTGGI
                YTTLNKTTQLPSTNKPTNSQAKSSTKPRVETHNKTTSNPAISLTDSADVPQRPREPTL
                PPIFRPPASKNRYLEKQLVIGLLTAVALTCGLITLFHYLFFR"
CDS             3194. .6592
                /note="ORF 06, major ssDNA binding protein homolog"
                /codon_start=1
                /protein_id="AAB62603.1"
                /db_xref="GI:2246478"
                /translation="MALKGPQTLEENIGSAAPTGPCGYLYAYLTHNFPIGEASLLGNG
                YPEAKVFSLPLLHGLTVESDFPLNVKAVHKKIDATTASVKLTSYHREAIVFHNTHLFQ
                PIFQGKGLEKLCRESRELFGFSTFVEQQHKGTLWSPEACPQLPCANEIFMAVIVTEGF
                KERLYGGKLVPVPSQTTPVHIGEHQAFKIPLYDEDLFGPSRAQELCRFYNPDISRYLH
                DSIFTGIAQALRVKDVSTVIQASERQFVHDQYKIPKLVQAKDFPQCASRGTDGSTLMV
                IDSLVAELGMSYGLSFIEGPQDSCEVLNYDTWPIFENCETPDARLRALEVWHAEQALH
                IGAQLFAANSVLYLTRVAKLPQKNQRGDANMYNSFYLQHGLGYLSEATVKENGASAFK
                GVPVSALDGSSYTLQHLAYASSFSPHLLARMCYYLQFLPHHKNTNSQSYNVVDYVGTA
                APSQMCDLCQGQCPAVCINTLFYRMKDRFPPVLSNVKRDPYVITGTAGTYNDLEILGN
                FATFREREEEGNPVEDAPKYTYWQLCQNITEKLASMGISEGGDALRTLIVDIPSFVKV
                FKGIDSTVEAELLKFINCMIKNNYNFRENIKSVHHILQFACNVYWQAPCPVFLTLYYK
                SLLTVIQDICLTSCMMYEQDNPAVGIVPSEWLKMHFQTMWTNFKGACFDKGAITGGEL
                KIVHQSMFCDLFDTDAAIGGMFAPARMQVRIARAMLMVPKTIKIKNRIIFSNSTGAES
                IQAGFMKPASQRDSYIVGGPYMKFLNALHKTLFPSTKTSALYLWHKIGQTTKNPILPG
                VSGEHLTELCNYVKASSQAFEEINVLDLVPDTLTSYAKIKLNSSILRACGQTQFYATT
                LSCLSPVTQLVPAEEYPHVLGPVGLSSPDEYRAKVAGRSVTIVQSTLKQAVSTNGRLR
                PIITVPLVVNKYTGSNGNTNVFHCANLGYFSGRGVDRNLRPESVPFKKNNVSSMLRKR
                HVIMTPLVDRLVKRIVGINSGEFEAEAVKRSVQNVLEDRDNPNLPKTVVLELVKHLGS
                SCASLTEEDVIYYLGPYAVLGDEVLSLLSTVGQAGVPWTAEGVASVIQDIIDDCELQF
                VGPEEPCLIQGQSVVEELFPSPGVPSLTVGKKRKIASLLSDLDL"
CDS             6609. .8696
                /note="ORF 7, processing and transport protein homolog
                (ICP18.5)"
                /codon_start=1
                /protein_id="AAB62620.1"
                /db_xref="GI:2246495"
                /translation="MAKELAAVYADVSALAMDLCLLSYADPATLDTKSLALTTGKFQS
                LHGTLLPLLRRQNAHECSGLSLELEHLLENVADALATLGVCTSRKLSPEEHFSLLHLD
                TTCNKHRSVRFNFYGNWALELKLSLINDVEIFFKRLSSVFYCIGSGSALEGLGEVLRF
                VGKLRGISPVPGPDLYVSNLPCLECLQEVCLTPNQGTSLQAMLPDTACSHICTPACGE
                PVRGLFENELKQLGLQTPESIPTTPCQSRVRQDDEIRQSSLMAVGDHHIFGEVTRSVL
                EISNLIYWSSGHSDATCDGDRDCSHLASLFTHEADMHKRRVDLAGCLGERGTPKHFFD
                CFRPDSLETLFCGGLFSSVEDTIESLQKDCSSAFYQQVNYTTALQKQNEFYVRLSKLL
                AAGQLNLGKCSTESCPSEARRQLVGGGKPEEVLRDAKHRQELYLQKVARDGFKKLSDC
                IRHQGHILSQTLGLRLWGSVIYNEASALQNHFLHRAQFISLPWQDLTVDCPTRFENSK
                YIKNSLYCQRLGREHVEILTLEFYKLITGPLSKRHTLFPSPPNVTLAQCFEAAGMLPH
                QKMMVSEMIWPSIEPKDWIEPNFNQFYSFENQDINHLQKRAWEYIRELVLSVSLYNRT
                WERELKILLTPQGSPGFEEPKPAGLTTGLYLTFETSAPLVLVDKKYGWIFKDLYALLY
                HHLQLSNHNDSQV"
CDS             8680. .11217
                /note="ORF 08, glycoprotein B (gB) homolog"
                /codon_start=1
                /protein_id="AAB62592.1"
                /db_xref="GI:2246467"
                /translation="MTPRSRLATLGTVILLVCFCAGAAHSRGDTFQTSSSPTPPGSSS
                KAPTKPGEEASGPKSVDFYQFRVCSASITGELFRFNLEQTCPDTKDKYHQEGILLVYK
                KNIVPHIFKVRRYRKIATSVTVYRGLTESAITNKYELPRPVPLYEISHMDSTYQCFSS
                MKVNVNGVENTFTDRDDVNTTVFLQPVEGLTDNIQRYFSQPVIYAEPGWFPGIYRVRT
                TVNCEIVDMIARSAEPYNYFVTSLGDTVEVSPFCYNESSCSTTPSNKNGLSVQVVLNH
                TVVTYSDRGTSPTPQNRIFVETGAYTLSWASESKTTAVCPLALWKTFPRSIQTTHEDS
                FHFVANEITATFTAPLTPVANFTDTYSCLTSDINTTLNASKAKLASTHVPNGTVQYFH
                TTGGLYLVWQPMSAINLTHAQGDSGNPTSSPPPSASPMTTSASRRKRRSASTAAAGGG
                GSTDNLSYTQLQFAYDKLRDGINQVLEELSRAWCREQVRDNLMWYELSKINPTSVMTA
                IYGRPVSAKFVGDAISVTECINVDQSSVNIHKSLRTNSKDVCYARPLVTFKFLNSSNL
                FTGQLGARNEIILTNNQVETCKDTCEHYFITRNETLVYKDYAYLRTINTTDISTLNTF
                IALNLSFIQNIDFKAIELYSSAEKRLASSVFDLETMFREYNYYTHRLAGLREDLDNTI
                DMNKERFVRDLSEIVADLGGIGKTVVNVASSVVTLCGSLVTGFINFIKHPLGGMLMII
                IVIAIILIIFMLSRRTNTIAQAPVKMIYPDVDRRAPPSGGAPTREEIKNILLGMHQLQ
                QEERQKADDLKKSTPSVFQRTANGLRQRLRGYKPLTQSLDISPETGE"
CDS             11344. .14382
                /note="ORF 09, DNA polymerase homolog"
                /codon_start=1
                /protein_id="AAB62593.1"
                /db_xref="GI:2246468"
                /translation="MDFFNPFIDPTRGGPRNTVRQPTPSQSPTVPSETRVCRLIPACF
                QTPGRPGVVAVDTTFPPTYFQGPKRGEVFAGETGSIWKTRRGQARNAPMSHLIFHVYD
                IVETTYTADRCEDVPFSFQTDIIPSGTVLKLLGRTLDGASVCVNVFRQRCYFYTLAPQ
                GVNLTHVLQQALQAGFGRASCGFSTEPVRKKILRAYDTQQYAVQKITLSSSPMMRTLS
                DRLTTCGCEVFESNVDAIRRFVLDHGFSTFGWYECSNPAPRTQARDSWTELEFDCSWE
                DLKFIPERTEWPPYTILSFDIECMGEKGFPNATQDEDMIIQISCVLHTVGNDKPYTRM
                LLGLGTCDPLPGVEVFEFPSEYDMLAAFLSMLRDYNVEFITGYNIANFDLPYIIARAT
                QVYDFKLQDFTKIKTGSVFEVHQPRGGSDGGNFMRSQSKVKISGIVPIDMYQVCREKL
                SLSDYKLDTVAKQCLGRQKDDISYKDIPPLFKSGPDGRAKVGNYCVIDSVLVMDLLLR
                FQTHVEISEIAKLAKIPTRRVLTDGQQIRVFSCLLEAAATEGYILPVPKGDAVSGYQG
                ATVISPSPGFYDDPVLVVDFASLYPSIIQAHNLCYSTLIPGDSLHLHPHLSPDDYETF
                VLSGGPVHFVKKHKRESLLAKLLTVWLAKRKEIRKTLASCTDPALKTILDKQQLAIKV
                TCNAVYGFTGVASGILPCLNIAETVTLQGRKMLERSQAFVEAISPERLAGLLRRPIDV
                SPDARFKVIYGDTDSLFICCMGFNMDSVSDFAEELASITTNTLFRSPIKLEAEKIFKC
                LLLLTKKRYVGVLSDDKVLMKGVDLIRKTACRFVQEKSSQVLDLILREPSVKAAAKLI
                SCQATDWVYREGLPEGFVKIIQVLNASHRELCERSVPVDKLTFTTELSRPLADYKTQN
                LPHLTVYQKLQARQEELPQIHDRIPYVFVDAPGSLRSELAEHPEYVKQHGLRVAVDLY
                FDKLVHAVANIIQCLFQNNTSATVAMLYNFLDIPVTFPTP"
CDS             14500. .15756
                /note="ORF 10, herpesvirus saimiri ORF 10 homolog,
                conserved in other gamma-herpesviruses"
                /codon_start=1
                /protein_id="AAB62594.1"
                /db_xref="GI:2246469"
                /translation="MQTEATFILGDWEITVSNCRFTCSSLTCGPLYRSSGDYTRLRIP
```

```
Query Match           18.4%;  Score 184.4;  DB 14;  Length 133661;
Best Local Similarity   57.8%;  Pred. No. 4.3e-33;
Matches  349;  Conservative    0;  Mismatches  251;  Indels    4;  Gaps    1;

Qy        112 GGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGAC 171
              ||   | | || ||| ||  | |       |||||  || |  |  |  | | |      |
Db      24828 GGGAGGGTGGCGCGCCGGGATCCCCCTCCGGGGGAGGGTGGCGCGCCGGGATCCCCCTCC 24769

Qy        172 AGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCT 231
              |  ||        |  |    |  |  |||||  ||| |    ||| |||||  |  |
Db      24768 GGGGGGGGGCGCGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGC 24709

Qy        232 AGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 291
                  ||| ||||  | || || || |  ||||  | || || || |  ||||  | || |
Db      24708 TCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGC 24649

Qy        292 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 351
              | || |  ||||  | || || || |  ||||  | || || || |  ||||  | || |
Db      24648 TCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGC 24589

Qy        352 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              | || |  ||||  | || || || |  ||||  | || || || |  ||||  | || |
Db      24588 TCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGC 24529

Qy        412 TGCT----TGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCC 467
              | ||    |||||  ||| |    ||| ||||  |  |       ||  |    |  | |
Db      24528 TCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGC 24469

Qy        468 TGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 527
              |  |||  || |    ||||| ||||| ||||    ||||| ||||| ||||    ||||
Db      24468 TCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGC 24409
```

```
Qy        528 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGC 587
              | ||||| ||||    ||||| ||||| ||||    ||||||||||| ||||    ||||
Db      24408 TCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGC 24349

Qy        588 TGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 647
              | ||||| ||||    |||||||||||| ||||    ||||| ||||| ||||    ||||
Db      24348 TCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGC 24289

Qy        648 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCT 707
              | ||||| ||||    ||||| ||||| ||||    ||||| |  |   ||     | ||
Db      24288 TCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCGCCGCTCCT 24229

Qy        708 TGCT 711
              | |
Db      24228 CGTT 24225
```

RESULT 37
AF013613

```
LOCUS       AF013613               28199 bp    DNA     linear   VRT 21-NOV-1997
DEFINITION  Fugu rubripes cosmid 48D10, complete sequence.
ACCESSION   AF013613
VERSION     AF013613.1  GI:2627437
KEYWORDS    HTG.
SOURCE      Takifugu rubripes
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 28199)
  AUTHORS   Sandford,R., Sgotto,B., Aparicio,S., Brenner,S., Vaudin,M.,
            Wilson,R., Chissoe,S., Pepin,K., Bateman,A., Chothia,C., Hughes,J.
            and Harris,P.
  TITLE     Comparative analysis of the polycystic kidney disease 1 (PKD1) gene
            reveals an integral membrane glycoprotein with multiple
            evolutionary conserved domains
  JOURNAL   Hum. Mol. Genet. 6 (9), 1483-1489 (1997)
  MEDLINE   97449170
   PUBMED   9285785
REFERENCE   2  (bases 1 to 28199)
  AUTHORS   Hawkins,J. and Gillam,B.
  TITLE     The sequence of Fugu rubripes 48D10
  JOURNAL   Unpublished (1997)
REFERENCE   3  (bases 1 to 28199)
  AUTHORS   Washington University Genome Sequencing Center.
  TITLE     A Fugu rubripes Genome Sequencing Project
  JOURNAL   Unpublished (1997)
REFERENCE   4  (bases 1 to 28199)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Nov 21, 1997 this sequence version replaced gi:2623547.
            Submitted by:
                    Genome Sequencing Center
                    Department of Genetics, Washington University,
                    St. Louis, MO 63108, USA
                    e-mail: rw@nematode.wustl.edu


            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone

            Library information:
            Gridded Fugu library from the UK HGMP Resource Centre at Hinxton.


                    NEIGHBORING COSMID INFORMATION:


            The left neighbor is 259C6, 19100 base overlap. The actual start of
            this clone is at base position 22192 of 259C6; the actual end is at
            base position 28199 of 48D10.

            This cosmid contains a tandem repeat from 23145 to 25326. The
            tandem is deleted, and has regions which are single stranded.
            Attempts to size the deletion with digest data were inconclusive.
            The size was estimated to be
            1.7 kb from a PCR product obtained by Dr. Dick Sandford.

            NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES             Location/Qualifiers
     source          1. .28199
                     /organism="Takifugu rubripes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:31033"
                     /chromosome="UL"
                     /map="unknown"
                     /clone="48D10"
     repeat_region   23145. .25326
                     /note="tandem with deletion"
     gene            26058. .28156
                     /gene="F_48D10.1"
     CDS             join(26058. .26702,27932. .28156)
                     /gene="F_48D10.1"
                     /note="Belongs to family 1 of G-protein coupled
                     receptors."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB86684.1"
                     /db_xref="GI:2627438"
                     /translation="MEPLDQTPGFPLSPEPNYWYETTPSLLLVSYPHLLDISSNQSTQ
                     SVPFQGSSALLTAVIYITVFVVGLTGNTLAIYVVLRYAGMKTVTNIYILNLAVADELY
                     IVGLPFLATQNVLSYWPFGSFLCRVVMTADSMNQFTSIFCLTVMSIDRYLAVVHPIRS
                     TKWRHPRVAKVVSAAVWAVSFVVVLPVVIFSDVQVRPSRPLQVGTSSKCLVKRVQETF
                     NSCNMIWPEPKNVWSTAFILYTAMVGFFGPLLIICLCYLLIVIKVRHRMSAAQVGAVV
                     STCPLNICCLSRR"
BASE COUNT     7492 a    6758 c    6389 g    7560 t
ORIGIN
```

```
  Query Match             18.3%;  Score 182.8;  DB 5;  Length 28199;
  Best Local Similarity   53.7%;  Pred. No. 1.1e-32;
  Matches  379;  Conservative    0;  Mismatches  327;  Indels    0;  Gaps    0;

Qy         41 GTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTG 100
              || ||| ||  |||  | |  |  |   | ||| |||||||   ||  | |||  ||
Db      23583 GTCTGCCTGTCTGTCTGTCTGCCTGCCTGTCTGTCTGCCTGTCTGCCTGTCTGTCTGTCT 23642

Qy        101 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 160
                | | || | |||| ||   |   | |||| |||      | | ||  | |||| |
Db      23643 GCCTGCCTGTCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGCCTGCCTGTCTGTCTGCCT 23702

Qy        161 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 220
              |       | | ||  |||||  ||   |   |||||| |||  |   | | ||  | ||
Db      23703 GTCTGTCTGCCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGCCTACCTGCCTGTCTGTCT 23762

Qy        221 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 280
              ||  |  |||  ||  | ||||  |   | |||  |||  ||  |   | |||| ||  |
Db      23763 GCCTGTCTGCCCGTCTGTCTGCCTGTCTGTCTGTCTGCCCGCCTGTCTGTCTGCCTGTCT 23822

Qy        281 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 340
              |   ||  | |||  ||| |||  ||  | |||| ||  ||   |   ||||  ||  |
Db      23823 GTCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGTCTGCCTGTCTGTCT 23882
```

```
Qy        341 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
              ||  ||  |||||  ||  |||  |   | ||||  |  ||   ||  | |||  ||| |
Db      23883 GCCTGCCTGCCTGTCTGTCTGCCTGTCTGTCTGCCCGTCTGTCTGCCTGTCTGTCTGCCT 23942

Qy        401 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTAC 460
              ||  |   | |||| ||        | | ||  | |||| ||| ||        | | ||
Db      23943 GCCTGTCTGTCTGCCTGTCTGTCTGCCTGTCTGTCTGCCTGCCTGTCTGTCTGTCTGTCT 24002

Qy        461 GCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 520
              |||||||||   || ||   ||  | ||||  | ||||  ||  | |||   | |||
Db      24003 GCCTGCCTGCCTGCCTGCCTGTCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGTCTGTCT 24062

Qy        521 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTA 580
              |   | ||||  | |||   |   | ||||  | ||||  |   | ||  |  | |||
Db      24063 GCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGTCT 24122

Qy        581 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 640
              ||  | ||||  | ||||  |   | ||    | ||||  ||  | ||||  | |||
Db      24123 GTCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTGCCTGTCTGTCTGCCTGTCTGTCTGTCT 24182

Qy        641 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTT 700
              ||  | ||||  | ||||  ||  | ||||  | ||||  |   | ||| |||   |  |
Db      24183 GTCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTGTCT 24242

Qy        701 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGC 746
              || ||| ||| |||  ||  | | || || ||||   |  | ||||
Db      24243 GCCTGCCTGCCTGCCTGTCTGTCTGCCTGTCTGCCTGTCTGTCTGC 24288


RESULT 38
AX652127
LOCUS       AX652127               28199 bp    DNA     linear   PAT 22-MAR-2003
DEFINITION  Sequence 1978 from Patent WO03000898.
ACCESSION   AX652127
VERSION     AX652127.1  GI:29154941
KEYWORDS    .
SOURCE      Takifugu rubripes (Fugu rubripes)
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Takifugu.
REFERENCE   1
  AUTHORS   Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
            Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
  TITLE     Plant genes involved in defense against pathogens
  JOURNAL   Patent: WO 03000898-A 1978 03-JAN-2003;
            Syngenta Participations AG (CH)
FEATURES             Location/Qualifiers
     source          1. .28199
                     /organism="Takifugu rubripes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:31033"
BASE COUNT     7492 a   6758 c   6389 g   7560 t
ORIGIN

  Query Match           18.3%;  Score 182.8;  DB 6;  Length 28199;
  Best Local Similarity   53.7%;  Pred. No. 1.1e-32;
  Matches  379;  Conservative    0;  Mismatches  327;  Indels    0;  Gaps    0;

Qy         41 GTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTG 100
              || ||| ||  |||  | |  |  |   | ||| |||||||   ||  | |||  ||
Db      23583 GTCTGCCTGTCTGTCTGTCTGCCTGCCTGTCTGTCTGCCTGTCTGCCTGTCTGTCTGTCT 23642

Qy        101 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 160
                | | ||  | |||| ||    |   | |||| |||      | | ||  | |||| |
Db      23643 GCCTGCCTGTCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGCCTGCCTGTCTGTCTGCCT 23702

Qy        161 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 220
              |       | | ||  |||||  ||   |  ||||| |||  |  | | ||  | ||
Db      23703 GTCTGTCTGCCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGCCTACCTGCCTGTCTGTCT 23762

Qy        221 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 280
              ||  |  |||  ||  | ||||  |   | |||  |||  ||  |   | ||||  ||  |
Db      23763 GCCTGTCTGCCCGTCTGTCTGCCTGTCTGTCTGTCTGCCCGCCTGTCTGTCTGCCTGTCT 23822

Qy        281 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 340
              |   ||  | |||  |||  |||  ||  | ||||  ||  ||   |  |||||  ||  |
Db      23823 GTCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGTCTGCCTGTCTGTCT 23882

Qy        341 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
              ||  ||  |||||  ||  |||  |   | ||||  |  ||   ||  | |||  ||| |
Db      23883 GCCTGCCTGCCTGTCTGTCTGCCTGTCTGTCTGCCCGTCTGTCTGCCTGTCTGTCTGCCT 23942

Qy        401 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTAC 460
              ||  |   | |||| ||        | | ||  | |||| ||| ||        | | ||
Db      23943 GCCTGTCTGTCTGCCTGTCTGTCTGCCTGTCTGTCTGCCTGCCTGTCTGTCTGTCTGTCT 24002

Qy        461 GCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 520
              |||||||||   || ||   ||  | ||||  | ||||  ||  | |||   | |||
Db      24003 GCCTGCCTGCCTGCCTGCCTGTCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGTCTGTCT 24062

Qy        521 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTA 580
              |   | ||||  | |||   |   | ||||  | ||||  |   | ||  |  | |||
Db      24063 GCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGTCT 24122

Qy        581 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 640
              ||  | ||||  | ||||  |   | ||    | ||||  ||  | ||||  | |||
Db      24123 GTCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTGCCTGTCTGTCTGCCTGTCTGTCTGTCT 24182

Qy        641 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTT 700
              ||  | ||||  | ||||  ||  | ||||  | ||||  |   | ||| |||   |  |
Db      24183 GTCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTGTCT 24242

Qy        701 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGC 746
              || ||| ||| |||  ||  | | || || ||||   |  | ||||
Db      24243 GCCTGCCTGCCTGCCTGTCTGTCTGCCTGTCTGCCTGTCTGTCTGC 24288


RESULT 39
AC109525_0
WPCOMMENT
  Sequence split into 5 fragments  LOCUS AC109525 Accession AC109525
     Fragment Name      Begin     End
     AC109525_0             1     110000
     AC109525_1        100001     210000
     AC109525_2        200001     310000
     AC109525_3        300001     410000
     AC109525_4        400001     478059
LOCUS       AC109525               478059 bp    DNA     linear   HTG 11-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-255D23, *** SEQUENCING IN PROGRESS
            ***, 5 unordered pieces.
ACCESSION   AC109525
VERSION     AC109525.6  GI:23603814
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 478059)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
```

```
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 478059)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 478059)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 9, 2002 this sequence version replaced gi:22795167.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GPHZ
                Center clone name: CH230-255D23
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 380245 bases at least Q40
                Consensus quality: 385911 bases at least Q30
                Consensus quality: 389757 bases at least Q20
                Estimated insert size: 390792; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   235023: contig of 235023 bp in length
            *    235024   235123: gap of unknown length
            *    235124   246065: contig of 10942 bp in length
            *    246066   246165: gap of unknown length
            *    246166   474601: contig of 228436 bp in length
            *    474602   474701: gap of unknown length
            *    474702   475761: contig of 1060 bp in length
            *    475762   475861: gap of unknown length
            *    475862   478059: contig of 2198 bp in length.
FEATURES             Location/Qualifiers
     source          1. .478059
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-255D23"
     misc_feature    1. .2769
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    complement(9835. .10732)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:MboI
                     end_sequence:RXACK24TVB"
     misc_feature    235124. .236699
                     /note="wgs_contig"
     misc_feature    287702. .289205
                     /note="wgs_contig"
     misc_feature    350281. .351366
                     /note="wgs_contig"
     misc_feature    358886. .361234
                     /note="wgs_contig"
     misc_feature    406035. .407045
                     /note="wgs_contig"
BASE COUNT   110790 a  84236 c  86247 g 111645 t  85141 others
ORIGIN

  Query Match            18.2%;  Score 182.4;  DB 2;  Length 110000;
  Best Local Similarity  57.6%;  Pred. No. 1.3e-32;
  Matches  348;  Conservative    0;  Mismatches  251;  Indels    5;  Gaps    1;

Qy        150 TAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 209
              | |||  | | ||||   ||   |  |||  ||| ||| |||||||  ||| ||||   ||
Db      52858 TTGGCTGGTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 52917

Qy        210 ACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 269
              || ||   | ||| | ||| || ||   | | | |||   | | | | | | |||
Db      52918 TTGGTTGGTTGGTTGGTTCGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGG 52977

Qy        270 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 329
              || ||| ||| |||||| ||| ||| ||| |||||| ||| ||| ||| | |||
Db      52978 TTGACTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGTTGG 53037

Qy        330 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 389
              |||| | | | | ||||||  ||| ||| | | | |||   ||| ||| ||| |||||
Db      53038 ATTGCTGGTTGGTTGGCTGGCTGGCTGGTTGGTTGGTTGGCTGGCTGGCTGGCTGG 53097

Qy        390 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAA 449
```

```
                 | | | | | | | |||   ||| ||||   ||    |  ||   ||| ||| |||
Db       53098 TTGGATGGTTGGTTGGTTGGTTGGCTGGCTGGCTGGTTGGCTGGCTGGCTGGCTGGCTGG 53157

Qy         450 GGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 509
               ||  |||     ||  || | | ||| |     || || | || || | |   || ||
Db       53158 CTGGCTGGT-----TGGTTGGTTGGTTGGTTCGCTGGTTGGTTGGTTGGTTGGTTGTTG 53212

Qy         510 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTC 569
               | || || | |   || || | || ||   |   || || | || || | |   || |
Db       53213 GTTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGTTG 53272

Qy         570 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTG 629
               |  |||| | |   || || | || || | |   || |  | || ||   |   |||||
Db       53273 GTTCGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGACTGGCTGGCTG 53332

Qy         630 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 689
               |||||   |   |||||  ||||| | | | || || | || || | |   |||||
Db       53333 GCTGGCTGGCTGGCTGGCTGGCTGGTTGGATGGTTGGTTGGTTGGTTGGCTGGCTG 53392

Qy         690 CCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGT 749
               ||| | |  || |||  || |||   |   ||| | |||  || |  |||||||   ||
Db       53393 GCTGGTTGGCTGGTTGGCTGGTTGGCTGGTTGGCTGGTTGGTTGATGATAGGGCTCAAGT 53452

Qy         750 GCAG 753
               | ||
Db       53453 GTAG 53456
```

```
RESULT 40
AC094443/c
LOCUS       AC094443               237019 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-4F6, *** SEQUENCING IN PROGRESS ***,
            3 unordered pieces.
ACCESSION   AC094443
VERSION     AC094443.6  GI:30466976
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 237019)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 237019)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 237019)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 9, 2003 this sequence version replaced gi:22772430.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GAPQ
                Center clone name: CH230-4F6
            -------------- Summary Statistics
                Assembly program: Atlas;
                Consensus quality: 229894 bases at least Q40
                Consensus quality: 231178 bases at least Q30
                Consensus quality: 232607 bases at least Q20
                Estimated insert size: 244941; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   231233: contig of 231233 bp in length
            *  231234   231333: gap of unknown length
```

```
          *   231334   233041: contig of 1708 bp in length
          *   233042   233141: gap of unknown length
          *   233142   237019: contig of 3878 bp in length.
FEATURES             Location/Qualifiers
     source          1. .237019
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-4F6"
     misc_feature    227374. .228903
                     /note="wgs_contig"
     misc_feature    228959. .231233
                     /note="wgs_contig"
BASE COUNT    73264 a  46066 c  45849 g  68157 t   3683 others
ORIGIN

  Query Match             18.2%;  Score 182.4;  DB 2;  Length 237019;
  Best Local Similarity   57.6%;  Pred. No. 1.2e-32;
  Matches  348;  Conservative    0;  Mismatches  251;  Indels    5;  Gaps    1;

Qy        150 TAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 209
              | |||  | | ||||   ||   |  |||  ||| ||| ||||||  ||| ||||   ||
Db      23724 TTGGCTGGTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 23665

Qy        210 ACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 269
              || ||   | ||| | ||| || ||   | | | |||   | | | | | | | |||
Db      23664 TTGGTTGGTTGGTTGGTTCGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGG 23605

Qy        270 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 329
              || ||| ||| ||||||  ||| ||| ||| ||||||  ||| ||| ||| | |||
Db      23604 TTGACTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGTTGG 23545

Qy        330 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 389
              |||| | | | | ||||||  ||| ||| | | | |||   ||| ||| ||| |||||
Db      23544 ATTGCTGGTTGGTTGGCTGGCTGGCTGGCTGGTTGGTTGGTTGGCTGGCTGGCTGGCTGG 23485

Qy        390 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAA 449
              | | | | | | | |||   ||| ||||   ||   |  ||   ||| ||| |||
Db      23484 TTGGATGGTTGGTTGGTTGGTTGGCTGGCTGGCTGGTTGGCTGGCTGGCTGGCTGGCTGG 23425

Qy        450 GGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 509
              ||  |||     ||  || | | ||| |     || || | || || | |   || ||
Db      23424 CTGGCTGGT-----TGGTTGGTTGGTTGGTTCGCTGGTTGGTTGGTTGGTTGGTTGGTTG 23370

Qy        510 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTC 569
              | || || | |   || || | || ||   |   || || | || || | |   || |
Db      23369 GTTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTG 23310

Qy        570 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTG 629
              |  |||| | |   || || | || || | |   || |  | || ||   |   |||||
Db      23309 GTTCGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGACTGGCTGGCTG 23250

Qy        630 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 689
              |||||   |   |||||  ||||| | | | || || | || || | |   |||||
Db      23249 GCTGGCTGGCTGGCTGGCTGGCTGGCTGGTTGGATGGTTGGTTGGTTGGTTGGCTGGCTG 23190

Qy        690 CCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGT 749
              ||| | |  || |||  || |||   |   ||| | |||  || |  |||||||   ||
Db      23189 GCTGGTTGGCTGGTTGGCTGGTTGGCTGGTTGGCTGGTTGGTTGATGATAGGGCTCAAGT 23130

Qy        750 GCAG 753
              | ||
Db      23129 GTAG 23126


RESULT 41
AC141889
LOCUS       AC141889              201258 bp    DNA     linear   HTG 20-MAR-2003
DEFINITION  Mus musculus chromosome UNK clone RP23-434P15, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC141889
VERSION     AC141889.2  GI:29126556
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 201258)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 201258)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 201258)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Mar 20, 2003 this sequence version replaced gi:29124296.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------- Project Information ----------
            Center project name: M_BA0434P15
            ----------
            ---------- Summary Statistics ----------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 199803 bases at least Q40
            Consensus quality: 200560 bases at least Q30
            Consensus quality: 201124 bases at least Q20
            ----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1386: contig of 1386 bp in length
            *     1387     1486: gap of unknown length
            *     1487     3048: contig of 1562 bp in length
            *     3049     3148: gap of unknown length
            *     3149     6869: contig of 3721 bp in length
            *     6870     6969: gap of unknown length
            *     6970    26526: contig of 19557 bp in length
            *    26527    26626: gap of unknown length
            *    26627    52028: contig of 25402 bp in length
            *    52029    52128: gap of unknown length
            *    52129    77727: contig of 25599 bp in length
            *    77728    77827: gap of unknown length
            *    77828   109907: contig of 32080 bp in length
            *   109908   110007: gap of unknown length
            *   110008   153565: contig of 43558 bp in length
            *   153566   153665: gap of unknown length
            *   153666   201258: contig of 47593 bp in length.
FEATURES             Location/Qualifiers
     source          1. .201258
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-434P15"
     misc_feature    1. .1386
                     /note="assembly_name:Contig17"
     misc_feature    1487. .3048
```

```
                     /note="assembly_name:Contig19"
     misc_feature    3149. .6869
                     /note="assembly_name:Contig20"
     misc_feature    6970. .26526
                     /note="assembly_name:Contig21"
     misc_feature    26627. .52028
                     /note="assembly_name:Contig22"
     misc_feature    52129. .77727
                     /note="assembly_name:Contig23"
     misc_feature    77828. .109907
                     /note="assembly_name:Contig24"
     misc_feature    110008. .153565
                     /note="assembly_name:Contig25"
     misc_feature    153666. .201258
                     /note="assembly_name:Contig26"
BASE COUNT    54579 a  44579 c  45083 g  56182 t    835 others
ORIGIN

  Query Match           18.0%;  Score 179.8;  DB 2;  Length 201258;
  Best Local Similarity   54.3%;  Pred. No. 5e-32;
  Matches  384;  Conservative    0;  Mismatches  322;  Indels    1;  Gaps    1;

Qy         64 GGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 123
Db      51174 GGTTGGTTGGCTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGCT 51233

Qy        124 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCT 183
Db      51234 GGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 51293

Qy        184 TGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCT 243
Db      51294 GGTTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGTT 51353

Qy        244 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 303
Db      51354 GGCTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGCTGGCTGGTTGG-TTGGTTGGTTGGT 51412

Qy        304 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 363
Db      51413 TGGTTGGTTGGTTGGTTGGCTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGT 51472

Qy        364 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGG 423
Db      51473 TGGCTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGC 51532

Qy        424 GGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTA 483
Db      51533 TGGCTGGTTGGCTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGCTGGT 51592

Qy        484 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 543
Db      51593 TGGCTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGCTGGCTGGTTGGT 51652

Qy        544 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 603
Db      51653 TGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGCTGGTTGGTTGGTTGGTTGGC 51712

Qy        604 GGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 663
Db      51713 TGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGT 51772

Qy        664 GGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGC 723
Db      51773 TGGTTGGTTGGTTGGTTGGTTGGCTGGCTGGTTGGCTGGTTGGCTGGTTGGTTGGCTGGT 51832

Qy        724 CGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAG 770
Db      51833 TGGTTGGCTGGTTGGTTGGCTGGTTGGCTGGTTGGTTGGCTGATTGG 51879


RESULT 42
AC091372
```

```
LOCUS       AC091372              235808 bp    DNA     linear   HTG 09-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-1C1, *** SEQUENCING IN PROGRESS ***,
            4 unordered pieces.
ACCESSION   AC091372
VERSION     AC091372.8  GI:22855787
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 235808)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 235808)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 235808)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 14, 2002 this sequence version replaced gi:22094155.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequening reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/).  As a result, the
            sequence may extend beyond the ends of the clone and there may be
            contigs that consist entirely of whole genome shotgun sequence
            reads.  Both end sequences and whole genome shotgun sequence only
```

```
          contigs will be indicated in the feature table.
          -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
              Center project name: TUFN
              Center clone name: CH230-1C1
          -------------- Summary Statistics
              Assembly program: Phrap; version 0.990329
              Consensus quality: 222117 bases at least Q40
              Consensus quality: 223886 bases at least Q30
              Consensus quality: 224946 bases at least Q20
              Estimated insert size: 242733; sum-of-contigs estimation
              Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 4 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1   232025: contig of 232025 bp in length
          *   232026   232125: gap of unknown length
          *   232126   233295: contig of 1170 bp in length
          *   233296   233395: gap of unknown length
          *   233396   234502: contig of 1107 bp in length
          *   234503   234602: gap of unknown length
          *   234603   235808: contig of 1206 bp in length.
FEATURES             Location/Qualifiers
     source          1. .235808
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-1C1"
     misc_feature    1. .1198
                     /note="wgs_contig"
     misc_feature    1541. .2900
                     /note="wgs_contig"
BASE COUNT    67252 a  46511 c  47166 g  65258 t   9621 others
ORIGIN

  Query Match             17.9%;  Score 178.6;  DB 2;  Length 235808;
  Best Local Similarity   55.0%;  Pred. No. 9.5e-32;
  Matches  394;  Conservative    0;  Mismatches  319;  Indels    4;  Gaps    2;

Qy          44 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 103
               ||| ||| ||| | |     |   | ||||||||||    ||  |||||| |||
Db      228837 TGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTTTC 228896

Qy         104 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 163
               | | ||  | |||| | |  ||  |||||  |||      | | |   |||||| |||
Db      228897 TGCCTGTCTGTCTGCCTACCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTGCCTGCCTGCC 228956

Qy         164 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 223
                 | | ||  | |||| |||  ||  |||||  |||      | | |     ||||
Db      228957 TTTCTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTACCTGCC 229016

Qy         224 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 283
               || |||  ||  | |||   |   |||||| ||| |||  ||  ||||   ||| ||
Db      229017 TGCCTGCCTGTCTGTCTGACTGACTGCCTGCCTGCCTGCCTGCCTTTCTGCCTGTC 229076

Qy         284 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 343
               |   |||||| ||| |||  |   |||||  ||  |||   |  |||||| ||| ||
Db      229077 TGTCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTACCTGCCTGCCTGCCTGTC 229136

Qy         344 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG---GCCTGCTTGCTT 400
               |   | |||| ||| |||  ||  |||||| ||  |||  ||||   |||||| |   |

Db      229137 TGTCTGACTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCTGCCTGCCTGCCTTTCT 229196

Qy         401 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTAC 460
               ||  |   | |||  |||      | | ||  | |||| ||  |||      | | ||
Db      229197 GCCTGTCTGTCTGACTGCCTGTCTGCCTGTCTGCCTGCCTGTCTGCCTGTCTGCCTGTCT 229256

Qy         461 GCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 520
               |||||||||   |  ||   ||  | ||||  | ||||  |   | ||||  | |||
Db      229257 GCCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGTCT 229316

Qy         521 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTA 580
               |     ||||  | ||||  ||  | |||   | ||||  |   | || |  | |||
Db      229317 GCCTACCTGCCTGCCTGCCTGTCTGTCTGACTGCCTGCCTGCCTGCCTGCCTGTCT 229376

Qy         581 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 640
               |    |||||  | ||||  |   | || |  | ||||  ||  | ||||  | |||
Db      229377 GCCT-GCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGTCT 229435

Qy         641 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTT 700
               |    ||||  | ||||  ||  | ||||  | ||||   |  | ||| |||   || |
Db      229436 GCCTATCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTTTCTGCCTGTCTGTCTGCCT 229495

Qy         701 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGG 757
               || ||  ||  |||  |   | | |  |||||||   |  | ||||| | ||   ||
Db      229496 GCCTGTCTGTCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTGCCTGCTATCCATGAGG 229552


RESULT 43
AC107437/c
LOCUS       AC107437              305226 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-84A6, *** SEQUENCING IN PROGRESS ***,
            25 unordered pieces.
ACCESSION   AC107437
VERSION     AC107437.6  GI:30580694
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 305226)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Huiyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
```

```
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 305226)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 305226)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:25188618.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GLDM
                Center clone name: CH230-84A6
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 283306 bases at least Q40
                Consensus quality: 287489 bases at least Q30
                Consensus quality: 290737 bases at least Q20
                Estimated insert size: 299662; sum-of-contigs estimation
                Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    11777: contig of 11777 bp in length
            *    11778    11877: gap of unknown length
            *    11878   258209: contig of 246332 bp in length
            *   258210   258309: gap of unknown length
            *   258310   259657: contig of 1348 bp in length
            *   259658   259757: gap of unknown length
            *   259758   261089: contig of 1332 bp in length
            *   261090   261189: gap of unknown length
            *   261190   262280: contig of 1091 bp in length
            *   262281   262380: gap of unknown length
            *   262381   264273: contig of 1893 bp in length
            *   264274   264373: gap of unknown length
            *   264374   265950: contig of 1577 bp in length
            *   265951   266050: gap of unknown length
            *   266051   267410: contig of 1360 bp in length
            *   267411   267510: gap of unknown length
            *   267511   268811: contig of 1301 bp in length
            *   268812   268911: gap of unknown length
            *   268912   270237: contig of 1326 bp in length
            *   270238   270337: gap of unknown length
            *   270338   271796: contig of 1459 bp in length
            *   271797   271896: gap of unknown length
            *   271897   274424: contig of 2528 bp in length
            *   274425   274524: gap of unknown length
            *   274525   275942: contig of 1418 bp in length
            *   275943   276042: gap of unknown length
            *   276043   277659: contig of 1617 bp in length
            *   277660   277759: gap of unknown length
            *   277760   279936: contig of 2177 bp in length
            *   279937   280036: gap of unknown length
            *   280037   281813: contig of 1777 bp in length
            *   281814   281913: gap of unknown length
            *   281914   283894: contig of 1981 bp in length
            *   283895   283994: gap of unknown length
            *   283995   286131: contig of 2137 bp in length
            *   286132   286231: gap of unknown length
            *   286232   287386: contig of 1155 bp in length
            *   287387   287486: gap of unknown length
            *   287487   289613: contig of 2127 bp in length
            *   289614   289713: gap of unknown length
            *   289714   291967: contig of 2254 bp in length
            *   291968   292067: gap of unknown length
            *   292068   293964: contig of 1897 bp in length
            *   293965   294064: gap of unknown length
            *   294065   295962: contig of 1898 bp in length
            *   295963   296062: gap of unknown length
            *   296063   300082: contig of 4020 bp in length
            *   300083   300182: gap of unknown length
            *   300183   305226: contig of 5044 bp in length.
FEATURES             Location/Qualifiers
     source          1. .305226
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-84A6"
     misc_feature    789. .37330
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH277447"
     misc_feature    10136. .11777
                     /note="wgs_contig"
     misc_feature    complement(239138. .240020)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH276631"
     misc_feature    253535. .254639
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    256559. .258209
                     /note="wgs_end_extension
                     clone_end:T7"
BASE COUNT    83055 a  60966 c  60899 g  88408 t  11898 others
ORIGIN
```

```
 Query Match             17.9%;  Score 178.6;  DB 2;  Length 305226;
 Best Local Similarity   55.0%;  Pred. No. 9.3e-32;
 Matches  394;  Conservative    0;  Mismatches  319;  Indels    4;  Gaps    2;

Qy         44 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 103
              ||| ||| |||  | |     |   | ||||||||||||   ||  |||||| |||
Db      88215 TGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTTTC 88156

Qy        104 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 163
              | | ||  | |||| | |  ||  |||||  |||       | | |   |||||| |||
Db      88155 TGCCTGTCTGTCTGCCTACCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTGCCTGCCTGCC 88096

Qy        164 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 223
                | | ||  | |||| |||  ||  |||||  |||       | | |     ||||
Db      88095 TTTCTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTACCTGCC 88036

Qy        224 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 283
              || |||  ||  | |||    |  ||||||  ||| |||  ||  ||||   ||| ||
Db      88035 TGCCTGCCTGTCTGTCTGACTGACTGCCTGCCTGCCTGCCTGCCTGCCTTTCTGCCTGTC 87976

Qy        284 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 343
              |   ||||||| ||| |||  |   |||||  ||  |||   |  |||||| ||| ||
Db      87975 TGTCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTACCTGCCTGCCTGCCTGTC 87916

Qy        344 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG---GCCTGCTTGCTT 400
              |   | |||| ||| |||  ||  |||||| ||  |||  ||||   |||||| |   |
Db      87915 TGTCTGACTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCTGCCTGCCTGCCTTTCT 87856

Qy        401 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTAC 460
              ||  |   | |||  |||       | | ||  | ||||  ||  |||       | | ||
Db      87855 GCCTGTCTGTCTGACTGCCTGTCTGCCTGTCTGCCTGCCTGTCTGCCTGTCTGCCTGTCT 87796

Qy        461 GCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 520
              |||||||||   |  ||   ||  | ||||  | ||||  |   | ||||  | |||
Db      87795 GCCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGTCT 87736

Qy        521 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTA 580
              |     ||||  | ||||  ||  | |||   | ||||  |   | ||  |  | |||
Db      87735 GCCTACCTGCCTGCCTGCCTGTCTGTCTGACTGCCTGCCTGCCTGCCTGCCTGCCTGTCT 87676

Qy        581 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 640
              |   ||||| | |||  |   | || |  | ||||  ||  | ||| |||  | |||
Db      87675 GCCT-GCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGTCT 87617

Qy        641 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTT 700
              |     ||||  | ||||  ||  | ||||  | ||||   |  | ||| |||   || |
Db      87616 GCCTATCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTTTCTGCCTGTCTGTCTGCCT 87557

Qy        701 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGG 757
              || ||  ||  |||  |   | | |  |||||||   |  | |||||  | ||   ||
Db      87556 GCCTGTCTGTCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTGCCTGCTATCCATGAGG 87500


RESULT 44
AC096597/c
LOCUS       AC096597               271233 bp    DNA     linear   HTG 13-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-21E2, *** SEQUENCING IN PROGRESS ***,
            5 unordered pieces.
ACCESSION   AC096597
VERSION     AC096597.6  GI:24941421
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 271233)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 271233)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 271233)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 13, 2002 this sequence version replaced gi:22772261.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ---------- Genome Center
                Center: Baylor College of Medicine
```

```
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
          ---------- Project Information
            Center project name: GEHU
            Center clone name: CH230-21E2
          ---------- Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 219869 bases at least Q40
            Consensus quality: 225773 bases at least Q30
            Consensus quality: 229378 bases at least Q20
            Estimated insert size: 221309; sum-of-contigs estimation
            Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
          ----------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 5 contigs. The true order of the pieces
          *.is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1    10670: contig of 10670 bp in length
          *    10671    10770: gap of unknown length
          *    10771   266157: contig of 255387 bp in length
          *   266158   266257: gap of unknown length
          *   266258   267281: contig of 1024 bp in length
          *   267282   267381: gap of unknown length
          *   267382   268591: contig of 1210 bp in length
          *   268592   268691: gap of unknown length
          *   268692   271233: contig of 2542 bp in length.
FEATURES             Location/Qualifiers
     source          1. .271233
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-21E2"
     misc_feature    6418. .8047
                     /note="wgs_contig"
     misc_feature    49552. .51724
                     /note="wgs_contig"
     misc_feature    148748. .149962
                     /note="wgs_contig"
     misc_feature    150054. .151277
                     /note="wgs_contig"
     misc_feature    247156. .248507
                     /note="wgs_contig"
BASE COUNT    68348 a  46760 c  47960 g  69593 t  38572 others
ORIGIN

  Query Match             17.8%;  Score 177.6;  DB 2;  Length 271233;
  Best Local Similarity   54.7%;  Pred. No. 1.6e-31;
  Matches  374;  Conservative    0;  Mismatches  309;  Indels    1;  Gaps    1;

Qy         64 GGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 123
              |||  |   |   ||  ||| |||||     | | ||   ||   || ||   | | | |
Db     221961 GGTTGGTTGGTTGGCTAGCTGGCTGGTTGTTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 221902

Qy        124 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCT 183
              ||||  |  ||| | ||   ||   || | |  | | | ||   ||   || ||   | |
Db     221901 GGCTGACTGGCTGGTTGGTTGGTTGGTTGACTGGTTGGTTGGTTGGTTGGTTGGATGGTT 221842

Qy        184 TGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCT 243
               ||| ||||||  ||| ||||   ||   || ||   | ||| | | | || ||   |||
Db     221841 GGCTGGCTGGCTGGCTGGCTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCT 221782

Qy        244 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 303
              ||| |  | | | | | | | || |  | | | | | | | | |||   | | | | |||
Db     221781 GACTGACTGGTTGGTTGGTTGGTTGACTGGTTGGTTGGTTGGTTGGTTGGTTGGCT 221722

Qy        304 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 363
              |||||| ||| ||| ||| | |||   | | | | | | |||||| ||| | | | |
Db     221721 GGCTGGCTGGCTGGCTGGCTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGGCTGGTTGGTT 221662

Qy         364 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGG 423
               | |||   | | | | |||  |||| || ||| ||| | | | | ||| ||| |  |
Db     221661 GGTTGGTTGGTTGGTTGGCTGACTGG-TTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGT 221603

Qy         424 GGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTA 483
               | | |   ||||| ||| ||| |  |  ||| |   |  ||  || | | ||| | |
Db     221602 TGGCTGACTGGCTGGTTGGTTGGTTGGTTGACTGGTTGGTTGGTTGGTTGGTTGGTTGGT 221543

Qy         484 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 543
               || ||   |||||   |   |||||  ||||| | |   || || | || || | |
Db     221542 TGGTTGGCTGGCTGGCTGGCTGGCTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGC 221483

Qy         544 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 603
               |||||  | ||| | |   || | |  | |||||   |   ||||| | || || | |
Db     221482 TGGCTGACTGACTGGTTGGTTGGTTGGTTGGCTGGCTGGCTGGCTGGTTGGTTGGTTGGC 221423

Qy         604 GGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 663
               ||||   | ||| | |   |||||  |||||   |   || || | || ||   |
Db     221422 TGGCTGGCTGACTGGTTGGCTGGCTGGCTGGCTGGCTGACTGGTTGGTTGGTTGGCTGGC 221363

Qy         664 GGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGC 723
               |||||  || || | |   |||||  || | | ||  ||| ||| |||   |   |||
Db     221362 TGGCTGACTGGTTGGTTGGTTGGCTGGTTGGTTGGTTCGTTGGTTGGTTGGCTGGCTGGC 221303

Qy         724 CGCTTGCCTGCTACTAGGGCTGCT 747
               | |||  |  |      || || |
Db     221302 TGGTTGGTTCGTTGGTTGGTTGGT 221279


RESULT 45
AF388526
LOCUS       AF388526                599 bp    DNA     linear   PRI 16-AUG-2001
DEFINITION  Homo sapiens ZNF9 gene, intron 1 and expanded CL3N58 repeat region.
ACCESSION   AF388526
VERSION     AF388526.1  GI:15193054
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 599)
  AUTHORS   Liquori,C.L., Ricker,K., Moseley,M.L., Jacobsen,J.F., Kress,W.,
            Naylor,S.L., Day,J.W. and Ranum,L.P.
  TITLE     Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of
            ZNF9
  JOURNAL   Science 293 (5531), 864-867 (2001)
  MEDLINE   21378719
   PUBMED   11486088
REFERENCE   2  (bases 1 to 599)
  AUTHORS   Liquori,C.L., Jacobsen,J.F. and Ranum,L.P.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-2001) Genetics, Cell Biology, and Development,
            University of Minnesota, 516 Delaware St. SE, Minneapolis, MN
            55455, USA
FEATURES             Location/Qualifiers
     source          1. .599
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3q21; between D3S1589 and D3S3606"
     gene            <1. .>599
                     /gene="ZNF9"
                     /note="synonym: DM2; myotonic dystrophy type 2"
     intron          <1. .>599
                     /gene="ZNF9"
                     /number=1
     repeat_region   1. .599
                     /note="CL3N58 repeat region"
     repeat_region   69. .100
```

```
                     /rpt_type=tandem
                     /rpt_unit=tg
     repeat_region   101. .136
                     /rpt_type=tandem
                     /rpt_unit=tctg
     repeat_region   137. .576
                     /note="expanded repeat in DM2 patients"
                     /rpt_type=tandem
                     /rpt_unit=cctg
BASE COUNT       22 a    248 c    156 g    173 t
ORIGIN
```

```
  Query Match             17.7%;  Score 177.4;  DB 9;  Length 599;
  Best Local Similarity   59.7%;  Pred. No. 2.8e-31;
  Matches  298;  Conservative    0;  Mismatches  201;  Indels    0;  Gaps    0;

Qy         39 GCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGC 98
              | ||||| |||  |||| | |     ||  | ||| ||| |||   |   | |||| |||
Db         82 GTGTGTGTGTGTGTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGCCTGC 141

Qy         99 TGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGC 158
                | | |   |||||| |||  ||  |||||| |||      | | |   ||||||
Db        142 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGC 201

Qy        159 TTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGG 218
               |||      | | |   |||||| ||| ||  |||||| |||      | | |   |
Db        202 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCC 261

Qy        219 CTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 278
              ||||  || |||  |   | ||||  ||  |||||| ||| |||  ||  |||||| |||
Db        262 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCC 321

Qy        279 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 338
               |||  ||  |||||| ||| |||  ||  |||||| ||| |||  ||  |||||| |||
Db        322 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCC 381

Qy        339 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 398
               |||  ||  |||||| ||| |||  ||  |||||| ||| |||  ||  |||||| |||
Db        382 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCC 441

Qy        399 TTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGT 458
               |||  ||  |||||| |||       | | |   | |||| ||| |||       | | |
Db        442 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGC 501

Qy        459 ACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 518
               |||||||||   || ||   |  | ||||  | ||||  |   | ||||  | ||||
Db        502 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCC 561

Qy        519 TAGTAGGGCTGCTGGGCTG 537
                |   | ||||  | |||
Db        562 CTGCCTGCCTGCCTGTCTG 580
```

```
RESULT 46
AC098644
LOCUS       AC098644               88716 bp    DNA     linear   HTG 30-AUG-2002
DEFINITION  Takifugu rubripes clone 205P11, WORKING DRAFT SEQUENCE, 7 ordered
            pieces.
ACCESSION   AC098644
VERSION     AC098644.3  GI:22549754
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Takifugu rubripes (Fugu rubripes)
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 88716)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
            Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Reddix-Dugue,N., Schueler,M.G., Sison,C.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 88716)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 88716)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-AUG-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Aug 30, 2002 this sequence version replaced gi:17155042.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: arp
                 Center clone name: 205P11

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 87141 bases at least Q40
                 Consensus quality: 87741 bases at least Q30
                 Consensus quality: 87993 bases at least Q20
                 Insert size: 89000; agarose-fp
                 Insert size: 88116; sum-of-contigs
                 Quality coverage: 10.68x in Q20 bases; agarose-fp
                 Quality coverage: 10.79x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    31156: contig of 31156 bp in length
            *    31157    31256: gap of unknown length
            *    31257    36515: contig of 5259 bp in length
            *    36516    36615: gap of unknown length
            *    36616    38675: contig of 2060 bp in length
            *    38676    38775: gap of unknown length
            *    38776    41147: contig of 2372 bp in length
            *    41148    41247: gap of unknown length
            *    41248    48183: contig of 6936 bp in length
            *    48184    48283: gap of unknown length
            *    48284    50318: contig of 2035 bp in length
            *    50319    50418: gap of unknown length
            *    50419    88716: contig of 38298 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1. .88716
                     /organism="Takifugu rubripes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:31033"
                     /clone="205P11"
                     /clone_lib="Incyte Genomics"
     misc_feature    1. .31156
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    31257. .36515
                     /note="assembly_fragment"
     misc_feature    36616. .38675
                     /note="assembly_fragment"
     misc_feature    38776. .41147
                     /note="assembly_fragment"
     misc_feature    41248. .48183
                     /note="assembly_fragment"
     misc_feature    48284. .50318
                     /note="assembly_fragment"
     misc_feature    50419. .88716
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    52511. .88716
                     /note="clone overlaps with GenBank Accession Number
                     AC098643 clone 194A9 (center project name aro)"
BASE COUNT    23629 a  19624 c  20682 g  24180 t    601 others
ORIGIN

  Query Match           17.7%;  Score 177.4;  DB 2;  Length 88716;
  Best Local Similarity  53.4%;  Pred. No. 1.9e-31;
  Matches  373;  Conservative    0;  Mismatches  326;  Indels    0;  Gaps    0;

Qy         48 TGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGAC 107
              ||  |||  | |  |  ||  | ||| ||| ||||  ||  |||||  |||      | |
Db      33801 TGTCTGTCTGTCTTCCTGTCTGTCTGTCTGTCTGCCTGCCTGTCTGCCTGTCTGCC 33860

Qy        108 AGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGG 167
              |   | |||| |||  |   ||||||| |||      | | ||  | |||  ||
Db      33861 TGCCTGTCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGTCTGTCTGTCTGTC 33920

Qy        168 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCT 227
              | |     | |||| ||   ||  |||||  ||       | | |   | |||   ||
Db      33921 TGCCTACCTGTCTGCCTGTCTGCCTGCCTGTCTGTCTGTCTGCCTGCCTGTCTGCC 33980

Qy        228 TGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 287
              |||  ||  | ||||  ||  |||||| ||| ||   ||  |||||| ||  ||   ||
Db      33981 TGCCTGTCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTGTCTGCC 34040

Qy        288 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 347
              | |||  ||| ||| ||  | |||| ||  ||   |   |||||| ||| || |  |
Db      34041 TGTCTGTCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGCCTGTTTACC 34100

Qy        348 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 407
              |||||  ||| |||  ||  | |||| ||  ||   ||  |||||  ||  ||   |
Db      34101 TGCCTGTCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTGCCTGTCTGTCTGTCTGTC 34160

Qy        408 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCC 467
              ||||||  || | |   | | ||  | |||| ||  ||       | | ||  | ||| |
Db      34161 TGCCTGCCTGTTTACCTGCCTGTCTGCCTGCCTGTCTGTCTGTCTGTCTGTCTGTC 34220

Qy        468 TGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 527
              ||   |  ||   |  | ||||  | ||||  ||  | |||   |  |||  ||  | |
Db      34221 TGTCTGTCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTGTCTGCATGCCTGTCTGTC 34280

Qy        528 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGC 587
              |||  | ||||  ||  | ||||  | |||   ||  | ||    | |||   |   | |
Db      34281 TGCCTGCCTGCCTGTCTGCCTGCCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGCCTGCC 34340

Qy        588 TGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 647
              |||  | ||||  |   | ||    | |||   ||  | ||||  | |||   ||  | |
Db      34341 TGCCTGCCTGCCTGCCTGTCTGTCTGTCTGTCTGTCTGTCTGCCTGCCTGTCTGTCTGTC 34400

Qy        648 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCT 707
              ||   | |||   ||  | ||||  | |||   ||  | ||||||||  |  ||| ||
Db      34401 TGTCTGTCTGTCTGTCTGCCTGCCTGCCTGTCTGTCTGTCTGCCTGCCTGTCTGCCTGTC 34460

Qy        708 TGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGC 746
              ||  |||  |   | | |  || |||    |  | ||||
Db      34461 TGTCTGCCTGCCTGTCTGTCTGTCTGTCTGTCTGCTGC 34499


RESULT 47
AC096407
LOCUS       AC096407              275443 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-59F13, *** SEQUENCING IN PROGRESS
            ***.
ACCESSION   AC096407
VERSION     AC096407.6  GI:30520472
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 275443)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
```

```
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 275443)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 275443)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24819680.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GFHU
                Center clone name: CH230-59F13
            -------------- Summary Statistics
               Assembly program: Atlas 3.0;
                Consensus quality: 261063 bases at least Q40
                Consensus quality: 264086 bases at least Q30
                Consensus quality: 266112 bases at least Q20
                Estimated insert size: 272725; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *         1   275443: contig of 275443 bp in length.
FEATURES                Location/Qualifiers
     source             1. .275443
                        /organism="Rattus norvegicus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10116"
                        /clone="CH230-59F13"
     misc_feature       1. .1557
                        /note="wgs_end_extension
                        clone_end:Sp6"
     misc_feature       2530. .2683
                        /note="clone_boundary
                        clone_end:Sp6
                        site:EcoRI
                        end_sequence:BH332249"
     misc_feature       complement(266627. .267346)
                        /note="clone_boundary
                        clone_end:T7
                        site:EcoRI
                        end_sequence:BH332247"
     misc_feature       272350. .274094
                        /note="wgs_end_extension
                        clone_end:T7"
     misc_feature       274145. .275443
                        /note="wgs_end_extension
                        clone_end:T7"
BASE COUNT    80036 a  54787 c  53947 g  78751 t   7922 others
ORIGIN

  Query Match             17.7%;  Score 177.4;  DB 2;  Length 275443;
  Best Local Similarity   55.0%;  Pred. No. 1.8e-31;
  Matches  369;  Conservative    0;  Mismatches  301;  Indels    1;  Gaps    1;

Qy         77 GCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTT 136
              ||  ||| |||||     | | ||   ||   || ||   | | | | |||| |  |||
Db      46377 GCTAGCTGGCTGGTTGTTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGACTGGCTG 46436

Qy        137 GCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 196
              | ||   ||   || | |  | | | ||   ||   || ||   | | ||| ||||||
Db      46437 GTTGGTTGGTTGGTTGACTGGTTGGTTGGTTGGTTGGTTGGATGGTTGGCTGGCTGGCTG 46496

Qy        197 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 256
              ||| ||||   ||  || ||   | ||| | | | || ||   |||  ||| |  | |
Db      46497 GCTGGCTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGACTGACTGGTTG 46556

Qy        257 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 316
              | | | | || |  | | | | | | |||   | | | | ||| ||||||  |||
Db      46557 GTTGGTTGGTTGACTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGGCTGGCTG 46616

Qy        317 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 376
              ||| ||| | |||   | | | | | | ||||||  ||| | | | | | |||   | |
Db      46617 GCTGGCTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGGCTGGTTGGTTGGTTGGTTGGTTG 46676

Qy        377 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 436
              | | |||  ||||  || ||| ||| | | | | ||| ||| |  |  | | |   ||||
Db      46677 GTTGGCTGACTGG-TTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGCTGACTGGCT 46735

Qy        437 GCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCT 496
              | ||| ||| |  |  ||| |   |  ||  || | | ||| | |   || ||   ||||
Db      46736 GGTTGGTTGGTTGGTTGACTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCT 46795

Qy        497 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 556
              |   |  |||||  ||||| | |   || || | || || | |   |||||   | ||
Db      46796 GGCTGGCTGGCTGGCTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGACTGACT 46855

Qy        557 GCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCT 616
              | | |   || | | |||||   |  ||||| | || || | |   ||||    | ||
Db      46856 GGTTGGTTGGTTGGTTGGCTGGCTGGCTGGCTGGTTGGTTGGTTGGCTGGCTGGCTGACT 46915

Qy        617 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 676
              | | |   |||||  |||||   |   || || | || ||   |  |||||  || |
Db      46916 GGTTGGCTGGCTGGCTGGCTGGCTGACTGGTTGGTTGGTTGGCTGGCTGGCTGACTGGTT 46975

Qy        677 GCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTA 736
              | | |  |||||  || | | ||  ||| ||| |||   |  ||| | |||  | |
Db      46976 GGTTGGTTGGCTGGTTGGTTGGTTCGTTGGTTGGTTGGCTGGCTGGCTGGTTGGTTCGTT 47035

Qy        737 CTAGGGCTGCT 747
                ||  || |
Db      47036 GGTTGGTTGGT 47046


RESULT 48
AC113967/c
LOCUS       AC113967              246968 bp    DNA     linear   HTG 21-OCT-2002
DEFINITION  Mus musculus clone RP23-59I22, WORKING DRAFT SEQUENCE, 3 ordered
            pieces.
ACCESSION   AC113967
VERSION     AC113967.4  GI:24182080
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
```

```
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 246968)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-59I22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 246968)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 246968)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 21, 2002 this sequence version replaced gi:21326293.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L23281
                Center clone name: 59_I_22
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 246357 bases at least Q40
                Consensus quality: 246703 bases at least Q30
                Consensus quality: 246763 bases at least Q20
                Insert size: 242000; agarose-fp
                Insert size: 246768; sum-of-contigs
                Quality coverage: 11.2 in Q20 bases; agarose-fp
                Quality coverage: 11.0 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    31803: contig of 31803 bp in length
            *   31804    31903: gap of 100 bp
            *   31904    80596: contig of 48693 bp in length
            *   80597    80696: gap of 100 bp
            *   80697   246968: contig of 166272 bp in length.
FEATURES             Location/Qualifiers
     source          1. .246968
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-59I22"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .31803
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    31904. .80596
                     /note="assembly_fragment"
     misc_feature    80697. .246968
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    73708 a  46388 c  45788 g  80884 t    200 others
ORIGIN

  Query Match             17.6%;  Score 176.4;  DB 2;  Length 246968;
  Best Local Similarity   53.1%;  Pred. No. 3.1e-31;
  Matches  398;  Conservative    0;  Mismatches  351;  Indels    1;  Gaps    1;

Qy          71 CTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGC 130
               |||||| |||||  |  |||| | |||     | |     |||| | |||   |  ||
Db      147996 CTGCCTCCCTGCCTCCCTGCCTCCCTGCCTCCCTGCCTCCCTGCCTCCCTGCCTCCCTGC 147937

Qy         131 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGC 190
               |||  |||      | | ||  |||||  |||      | | ||  |||||  |||  |
Db      147936 CTGTCTGCCTCCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGT 147877

Qy         191 TGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGC 250
                 |||||  |||      | | ||  | ||||   | |||  ||  | | || |  ||  |
Db      147876 CTGCCTGTCTGCCTCCCTGCCTGTCTGTCTGCCTCCCTGCCTGTCTGCCTCCCTGCCTGT 147817

Qy         251 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 310
               |||| | | |||  |   |||| | ||| ||   ||  | |||| ||  |||  |   ||
Db      147816 CTGCCTCCCTGCCTGTCTGCCTCCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGC 147757

Qy         311 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 370
               |||  ||| ||   ||  | |||| ||  |||  |   |||||  |||  ||   ||  ||
Db      147756 CTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGC 147697

Qy         371 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 430
               |||| ||| |||  ||  |||||| ||| |||  ||  |||||| |||       | | |
Db      147696 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGC 147637

Qy         431 AGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGC 490
                 | |||| ||| |||      | | |   | |||||||   || ||   |   | ||||
Db      147636 CTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGC 147577
```

```
Qy        491 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 550
Db     147576 CTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGC 147517

Qy        551 TGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCC 610
Db     147516 CTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTG- 147458

Qy        611 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 670
Db     147457 TCGTCTGTCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGT 147398

Qy        671 TGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGC 730
Db     147397 CTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGC 147338

Qy        731 CTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGGGCTGCCGGCCAGCTGCGTG 790
Db     147337 CTGTCTGTCTGCCTGTCATCTGCCTGCCTGCAGATTCTGACTGACTCTGTTAACTTCCTG 147278

Qy        791 CGAGGGCGTCCGAGGGCCAGACGAGGACAC 820
Db     147277 CCAGATGATCCATATCCCTAACTGGGCCAC 147248


RESULT 49
AC141889/c
LOCUS       AC141889              201258 bp    DNA     linear   HTG 20-MAR-2003
DEFINITION  Mus musculus chromosome UNK clone RP23-434P15, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC141889
VERSION     AC141889.2  GI:29126556
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 201258)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 201258)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 201258)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Mar 20, 2003 this sequence version replaced gi:29124296.

            ---------------- Genome Center ----------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information -----------------
            Center project name: M_BA0434P15
            ----------------------------------------------------
            -------------- Summary Statistics ------------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 199803 bases at least Q40
            Consensus quality: 200560 bases at least Q30
            Consensus quality: 201124 bases at least Q20
            ----------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1386: contig of 1386 bp in length
            *     1387     1486: gap of unknown length
            *     1487     3048: contig of 1562 bp in length
            *     3049     3148: gap of unknown length
            *     3149     6869: contig of 3721 bp in length
            *     6870     6969: gap of unknown length
            *     6970    26526: contig of 19557 bp in length
            *    26527    26626: gap of unknown length
            *    26627    52028: contig of 25402 bp in length
            *    52029    52128: gap of unknown length
            *    52129    77727: contig of 25599 bp in length
            *    77728    77827: gap of unknown length
            *    77828   109907: contig of 32080 bp in length
            *   109908   110007: gap of unknown length
            *   110008   153565: contig of 43558 bp in length
            *   153566   153665: gap of unknown length
            *   153666   201258: contig of 47593 bp in length.
FEATURES             Location/Qualifiers
     source          1. .201258
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-434P15"
     misc_feature    1. .1386
                     /note="assembly_name:Contig17"
     misc_feature    1487. .3048
                     /note="assembly_name:Contig19"
     misc_feature    3149. .6869
                     /note="assembly_name:Contig20"
     misc_feature    6970. .26526
                     /note="assembly_name:Contig21"
     misc_feature    26627. .52028
                     /note="assembly_name:Contig22"
     misc_feature    52129. .77727
                     /note="assembly_name:Contig23"
     misc_feature    77828. .109907
                     /note="assembly_name:Contig24"
     misc_feature    110008. .153565
                     /note="assembly_name:Contig25"
     misc_feature    153666. .201258
                     /note="assembly_name:Contig26"
BASE COUNT    54579 a  44579 c  45083 g  56182 t    835 others
ORIGIN

  Query Match            17.5%;  Score 174.8;  DB 2;  Length 201258;
  Best Local Similarity  54.2%;  Pred. No. 7.5e-31;
  Matches  378;  Conservative    0;  Mismatches  317;  Indels    3;  Gaps    1;

Qy         64 GGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 123
Db      76525 GGTTGGCTGGTTTGTTGGCTGGCTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGCTGGCT 76466

Qy        124 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCT 183
Db      76465 GGCTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCT 76406

Qy        184 TGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCT 243
Db      76405 GGCTGGCTGGTTGGCTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGCTGGCTGGCTGGTT 76346

Qy        244 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC---CTGCTTGCTT 300
Db      76345 GGTTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGGCTGGCTGGCTGGTTGGTT 76286

Qy        301 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 360
```

```
Db      76285 GGTTGGTTGGTTGGCTGGTTGATTGGCTGGTTGGTTGGCTGGTTGGTTGGTTGGTT 76226

Qy        361 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 420

Db      76225 GGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTT 76166

Qy        421 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 480

Db      76165 GGTTGGTTGGTTAGTTGGTTGGTTGGTTGGTTGGCTGGTAGGTTGGTTGGTTGGTTGGCT 76106

Qy        481 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 540

Db      76105 GGTTGGCTGGTTGGCTAAAAGGTTGGTTGGTTGGCTGGTTGGTTAGTTGGTTGGATGGTT 76046

Qy        541 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 600

Db      76045 GGTTGGTTGGTTGGCTGGTTGGTTGGTTGGCTAGTTAGCTAGTTGGCTGGTTGGTTGGTT 75986

Qy        601 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 660

Db      75985 GGTTAGCTGGTTGGCTGATTGGTTGGTTGATTGGTTGGTTGACTGGTTGGTTGGCTGGTT 75926

Qy        661 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGG 720

Db      75925 GGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 75866

Qy        721 GGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGG 758

Db      75865 GGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGG 75828
```

RESULT 50
AC115149
LOCUS AC115149 242482 bp DNA linear HTG 13-NOV-2002
DEFINITION Rattus norvegicus clone CH230-95C4, WORKING DRAFT SEQUENCE.
ACCESSION AC115149
VERSION AC115149.4 GI:24941931
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 242482)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 242482)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (15-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 242482)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 13, 2002 this sequence version replaced gi:23665230.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GOPR
Center clone name: CH230-95C4
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 222780 bases at least Q40
Consensus quality: 224645 bases at least Q30
Consensus quality: 225630 bases at least Q20
Estimated insert size: 230869; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and

```
          * the accession number will be preserved.
          *        1   242482: contig of 242482 bp in length.
FEATURES             Location/Qualifiers
     source          1. .242482
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-95C4"
     misc_feature    1. .1190
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    1692. .3431
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    4820. .5678
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BH357978"
BASE COUNT    56337 a  54961 c  55419 g  59910 t  15855 others
ORIGIN

  Query Match             17.3%;  Score 173.4;  DB 2;  Length 242482;
  Best Local Similarity   53.1%;  Pred. No. 1.6e-30;
  Matches  369;  Conservative    0;  Mismatches  326;  Indels    0;  Gaps    0;

Qy         41 GTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTG 100
              |||||   |  ||   |      |||  |   || ||   |   |||||  || | ||
Db     217930 GTGTGGTGGTGTGCTGGCGTGGTGGTGTGCTGGCGTGGTGGTGTGCTGGTGTGGTGGCAT 217989

Qy        101 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 160
              | ||   |  ||| || | |   ||||| || | ||  |    |  | ||  ||||
Db     217990 GGTGGTGTGCTGGCGTGGTGGTGTGCTGGCGTGGTGGCGTAAAAAAGTGGTGGTGTGCTG 218049

Qy        161 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 220
              ||   | ||   |  ||  || | |   |||||  || | ||   | ||   |  ||  |
Db     218050 GCGTGGTGGTGTGCTGGTGTGGTGGTGTGCTGGTGTGGTGGCATGGTGGTGTGCTGGCGT 218109

Qy        221 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 280
              | |||| || | |   ||  ||  |||||| || | || || | ||  |   || || |
Db     218110 GGTGGCGTGGTGGCGTGGTGGCGTGCTGGCGTGGTGGCGTGGTGGCGTGGTGGCGTGGTG 218169

Qy        281 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 340
              ||  | |||| || | || |||| ||  |   || |||| ||  | |||  |||| || |
Db     218170 GCGTGGTGGCGTGGTGGCGTGCTGGCGTGGTGGCGTGCTGGCGTGGTGGTGTGCTGGCGT 218229

Qy        341 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
              | | |   ||  || || | |   |||||| || | || |||| ||  |   || || |
Db     218230 GGTGGTGTGCTGGCGTGGTGGTGTGCTGGCGTGGTGGCGTGCTGGCGTGGTGGCGTGGTG 218289

Qy        401 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTAC 460
              ||  | |||| |||| ||   | ||   |  ||  |||| || || |   | |  ||
Db     218290 GCGTGGTGGCGTGCTGGCGTGGTGGTGTGCTGGCGTGCTGGCGTGGTGGCGTGCTGGCGT 218349

Qy        461 GCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 520
              |   |  || |||| || | ||  ||  ||  ||  |   |  ||  || |||  ||||
Db     218350 GGTGGTGTGCTGGCGTGGTGGTGTGGTGGCGTGGTGGTGTGCTGGTGTGGTGGTGTGCTG 218409

Qy        521 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTA 580
              ||  ||  ||  ||  |   |  ||  || |||  || | |   ||   |  ||  ||
Db     218410 GTGTGGTGGCATGGTGGTGTGCTGGTGTGGTGGCATGGTGGCGTGGTGGCGTGGTGGCGT 218469

Qy        581 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 640
              |  ||  || |||  ||||  |   ||   |  |   ||  |  ||  ||||||  || |
Db     218470 GCTGGCGTGGTGGCGTGCTGGCGTGGTGGCGTGCTGGCGTGGTGGCGTGCTGGCGTGGTG 218529

Qy        641 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTT 700
              ||  |   ||  ||  |   |  ||  || |||  ||||  |   ||   |  |||||| |
Db     218530 GTGTGCTGGCGTGGTGGTGTGCTGGCGTGGTGGTGTGCTGGCGTGGTGGTGTGCTGGCGT 218589

Qy        701 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCT 735
              | | || |||| ||  |  ||   | | |  || |
Db     218590 GGTGGCGTGCTGGCGTGGTGGCGTGGTGGTGTGGT 218624
```

```
RESULT 51
AC113840/c
LOCUS       AC113840               245258 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-126F2, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC113840
VERSION     AC113840.5  GI:25072829
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 245258)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 245258)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
```

```
REFERENCE   3  (bases 1 to 245258)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:23101014.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GSOU
                Center clone name: CH230-126F2
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 228855 bases at least Q40
                Consensus quality: 231441 bases at least Q30
                Consensus quality: 233224 bases at least Q20
                Estimated insert size: 235507; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    78039: contig of 78039 bp in length
            *    78040    78139: gap of unknown length
            *    78140   241585: contig of 163446 bp in length
            *   241586   241685: gap of unknown length
            *   241686   245258: contig of 3573 bp in length.
FEATURES             Location/Qualifiers
     source          1. .245258
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-126F2"
     misc_feature    1683. .2513
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BH358726"
     misc_feature    6441. .7279
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BH358725"
     misc_feature    78140. .80241
                     /note="wgs_end_extension
                     clone_end:T7"
BASE COUNT    62337 a  53440 c  55280 g  63518 t  10683 others
ORIGIN

  Query Match             17.1%;  Score 171;  DB 2;  Length 245258;
  Best Local Similarity   53.6%;  Pred. No. 5.7e-30;
  Matches  377;  Conservative    0;  Mismatches  325;  Indels    1;  Gaps    1;

Qy         44 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 103
              ||  ||| ||   | |  |   |  | ||||||||||||  |   |||||  | |
Db      98828 TGTCTGCCTGCCTGTCTGCCTATCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTCCCTGCC 98769

Qy        104 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 163
              | | |     |||| ||| ||  |||||| ||       | | |   ||||   |||
Db      98768 TGCCTGCCTATCTGCCTGCCTGCCTGCCTGCCTGTCTCCCTGCCTGCCTGCCTATCTGCC 98709

Qy        164 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 223
              | | ||  |||||| |||  ||  |||||  |||      | | ||  | ||||
Db      98708 TGTCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGTCTGCC 98649

Qy        224 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 283
              || |||  ||  | |||   |    |||| ||  ||   ||  | |||| ||| ||
Db      98648 TGCCTGCCTGTCTGTCTGACTGTCTACCTGTCTGTCTGTCTGCCTGTCTGCCTGCCTGTC 98589

Qy        284 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 343
              |   | ||  |  ||  ||| ||  |||||  ||| |||  |    |||| ||| |||
Db      98588 TGACTGTCTACCTGTCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTCCCTGCCTGCCTGCC 98529

Qy        344 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403
              |||||  ||| ||   ||  |||||| ||| || ||    |||||  ||| |||
Db      98528 TATCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTG-TCTGCCGCCTGTCTGCCTGCC 98470

Qy        404 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCC 463
              |   |||||| |||  |   | | |      |||| |   |||  |   | | ||  |||
Db      98469 TGTCTGCCTGCCTGCCTACCTGTCTGCCTATCTGCCTATCTGCCTATCTGCCTGTCTGCC 98410

Qy        464 TGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 523
              ||||||   || |    |   | ||||  | |||   |   | ||||  | ||||  ||
Db      98409 TGCCTGCCTGCCTATCTGCCTGCCTGCCTGTCTGTCTGACTGTCTGCCTGTCTGCCTGTC 98350

Qy        524 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTA 583
              | ||||  | ||||  ||  | ||||  | |||   |   | ||    | ||||  ||
Db      98349 TGTCTGCCTGTCTGCCTGTCTGTCTGCCTGTCTGTCTGCCTGCCTGTCTGTCTGCCTGTC 98290

Qy        584 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 643
              | ||||  | |||       | || |  | ||||   |    |||  | |||   ||
Db      98289 TGCCTGCCTGCCTGTCTCCCTGCCTGCCTGCCTGCCTATCTTCCTGTCTGCCTGTCTGTC 98230

Qy        644 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCT 703
              | ||||    ||||  ||  | ||||  | ||||   |  | ||| ||||  |  |||
Db      98229 TGCCTGCCTATCTGCCTGTCTGTCTGCCTGCCTGCCTATCTGCCTGTCTGCCTGTCTGCC 98170

Qy        704 TGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGC 746
              ||  ||  |||  ||  | | || ||||||       | ||||
Db      98169 TGTCTGTCTGCCTGTCTGCCTGCCTGCCTGTCTGCCTGTCTGC 98127


RESULT 52
AL845305
LOCUS       AL845305              200141 bp    DNA     linear   ROD 02-DEC-2002
DEFINITION  Mouse DNA sequence from clone RP23-202K13 on chromosome 2, complete
            sequence.
ACCESSION   AL845305
VERSION     AL845305.5  GI:26185503
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 200141)
  AUTHORS   Smith,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 6, 2002 this sequence version replaced gi:23573260.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to confirm this sequence.  Sequence data
```

from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.
---------- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-202K13 is from the RPCI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.

```
FEATURES             Location/Qualifiers
     source          1. .200141
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-202K13"
                     /clone_lib="RPCI-23"
BASE COUNT    63702 a  37575 c  38312 g  60552 t
ORIGIN

  Query Match             17.1%;  Score 170.6;  DB 10;  Length 200141;
  Best Local Similarity   67.0%;  Pred. No. 7.2e-30;
  Matches  242;  Conservative    0;  Mismatches  119;  Indels    0;  Gaps    0;

Qy          85 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 144
               |||   ||  ||  |||  ||   |   || |||||||  ||| ||||||||||
Db      185420 GCTTTGCTTTTTAATGAAAGGCTTGCTTGCTTGCTTGCCTGCTTGCCTGCTTGCCTGCCT 185479

Qy          145 GACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 204
               | | |   || ||| ||||     |   |   || ||| |||| ||| || ||||||||
Db      185480 GCCTGCCTGCTTGCCTGCTTGTCTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTGCTT 185539

Qy          205 AGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTC 264
                   |   |   | |||| ||||||| |   | ||||  ||  |||||| ||| |||
Db      185540 GCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 185599

Qy          265 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 324
               ||  |||||| ||||||   ||| ||||||||||| |||| ||| || ||||||||||
Db      185600 GCCTGCCTGCCTGCTTGTCTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCCT 185659

Qy          325 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 384
               ||| ||||||||||| |||| ||| || ||||||||||||| ||| || ||||||||||||
Db      185660 GCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 185719

Qy          385 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 444
               ||| || |||||||||| |||||  || ||||||    |  |   | | || ||  ||  |
Db      185720 GCTTGCTTGCTTGCTTCCTCGCCTGCTTGCTTGGATGGATGGATGGATGGATGGATGGAT 185779

Qy          445 G 445
                |
Db      185780 G 185780
```

RESULT 53
AC118832

```
LOCUS       AC118832               206805 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-271G9, *** SEQUENCING IN PROGRESS
            ***, 12 unordered pieces.
ACCESSION   AC118832
VERSION     AC118832.4  GI:25094002
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 206805)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 206805)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
```

```
REFERENCE   3  (bases 1 to 206805)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:23908265.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GVTR
                Center clone name: CH230-271G9
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 172120 bases at least Q40
                Consensus quality: 175720 bases at least Q30
                Consensus quality: 178116 bases at least Q20
                Estimated insert size: 177937; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    48963: contig of 48963 bp in length
            *    48964    49063: gap of unknown length
            *    49064    77623: contig of 28560 bp in length
            *    77624    77723: gap of unknown length
            *    77724    88733: contig of 11010 bp in length
            *    88734    88833: gap of unknown length
            *    88834   169408: contig of 80575 bp in length
            *   169409   169508: gap of unknown length
            *   169509   180885: contig of 11377 bp in length
            *   180886   180985: gap of unknown length
            *   180986   191517: contig of 10532 bp in length
            *   191518   191617: gap of unknown length
            *   191618   193312: contig of 1695 bp in length
            *   193313   193412: gap of unknown length
            *   193413   194460: contig of 1048 bp in length
            *   194461   194560: gap of unknown length
            *   194561   195655: contig of 1095 bp in length
            *   195656   195755: gap of unknown length
            *   195756   198673: contig of 2918 bp in length
            *   198674   198773: gap of unknown length
            *   198774   202107: contig of 3334 bp in length
            *   202108   202207: gap of unknown length
            *   202208   206805: contig of 4598 bp in length.
FEATURES             Location/Qualifiers
     source          1. .206805
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-271G9"
     misc_feature    1. .1288
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    complement(9131. .9885)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BZ218662"
     misc_feature    164162. .165223
                     /note="wgs_contig"
     misc_feature    169509. .171921
                     /note="wgs_contig"
     misc_feature    180986. .182266
                     /note="wgs_contig"
     misc_feature    185783. .187052
                     /note="wgs_contig"
BASE COUNT    53759 a  37440 c  37301 g  51783 t  26522 others
ORIGIN

  Query Match             17.1%;  Score 170.6;  DB 2;  Length 206805;
  Best Local Similarity   54.8%;  Pred. No. 7.2e-30;
  Matches  355;  Conservative    0;  Mismatches  292;  Indels    1;  Gaps    1;

Qy          81 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 140
               ||| | |||| | | | ||  ||   || ||| | | | | |||||   ||| | ||
Db      154265 GCTGGTTGGCTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGCTGGTTG 154324

Qy         141 AGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTT 200
                 ||   || | |  ||| ||||   ||   || ||   | | | | | ||||  | |
Db      154325 GTTGGTTGGTTGACTGGCTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGTTG 154384

Qy         201 GCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTT 260
               ||||   ||   |  ||   ||||||| | | || ||  | | | |||   ||| | |
Db      154385 GCTGGTTGGTTGGCTGGTTGGCTGGCTGGTTGGTTGGTTGGTTGGTTGGCTGGTTG 154444

Qy         261 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 320
               | | |||||  || ||| ||| | | | |  || ||| ||| | |   |  || ||| ||
Db      154445 GTTGGCTGG-TTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTT 154503

Qy         321 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 380
               | | | | |  ||  ||  || | | | |  || ||| ||| | | | | ||| ||| ||
Db      154504 GGTTGGTTGGTTGACTGACTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTT 154563

Qy         381 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTT 440
               |   | | | ||| ||| ||| | | | | |||  || | |  |  |   || || ||
Db      154564 GGCTGGTTGGCTGGTTGGTTGGTTGGTTGACTGACTGGTTGGTTGGTTGGTTGGTT 154623

Qy         441 GCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTA 500
               | ||| | |  |   |  || ||| | | ||| | |   ||||| | ||||| |
Db      154624 GGTTGGTTGGCTGGTTGGCTGGTTGGCTGGTTGGTTGGTTGGTTGGCTGGTTGGCTGGTT 154683

Qy         501 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 560
               |  || ||  |  |  | |    |||| | || || | ||  || || | ||||| |
Db      154684 GGCTGGTTGGCTGTTTTGTTGATTCGCTGGTTGGTTGGTTGTTTGGTTGGTTGGCTGGTT 154743

Qy         561 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTA 620
               |  |  |  | ||||  |    |  |  | ||||  | ||  || |  | | ||| |
Db      154744 GGTTGTTTGGTTGGCTAGTTNGTTGCNTTGTTGGCTCGTTGTTTGGTTGGTAGNCTGGTT 154803

Qy         621 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 680
               |  ||      || || | |  ||||       || | ||  ||||| | ||||| ||
Db      154804 GGTTGGTCCGCTGGTTGGTTGGTTGGCTGGCTCCTTGGTCGTCTGGCTGGTTGGCTGGTA 154863

Qy         681 GTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTAGTGGGGCCGCTT 728
               || ||| |  ||| |  |||| ||   || ||| | |   ||  | ||
Db      154864 GTTGGGATAGCTGGTTTCTTGATTTGCTGATTGGTTGGTTGGTTGGTT 154911


RESULT 54
AC116951
LOCUS       AC116951              256110 bp    DNA     linear   HTG 01-SEP-2002
DEFINITION  Mus musculus chromosome UNK clone RP23-75O6, WORKING DRAFT
```

```
            SEQUENCE, 12 unordered pieces.
ACCESSION   AC116951
VERSION     AC116951.2  GI:22539388
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 256110)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 256110)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 256110)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Aug 29, 2002 this sequence version replaced gi:19909483.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------- Project Information ----------
            Center project name: M_BA0075O06
            ----------
            ---------- Summary Statistics ----------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 252203 bases at least Q40
            Consensus quality: 253256 bases at least Q30
            Consensus quality: 254009 bases at least Q20
            Insert size: 216000; agarose-fp
            Insert size: 261658; sum-of-contigs
            Quality coverage: 13.71 in Q20 bases; agarose-fp
            Quality coverage: 9.96 in Q20 bases; sum-of-contigs
            ----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1235: contig of 1235 bp in length
            *     1236     1335: gap of unknown length
            *     1336     3639: contig of 2304 bp in length
            *     3640     3739: gap of unknown length
            *     3740     5766: contig of 2027 bp in length
            *     5767     5866: gap of unknown length
            *     5867     8394: contig of 2528 bp in length
            *     8395     8494: gap of unknown length
            *     8495    12007: contig of 3513 bp in length
            *    12008    12107: gap of unknown length
            *    12108    17198: contig of 5091 bp in length
            *    17199    17298: gap of unknown length
            *    17299    24643: contig of 7345 bp in length
            *    24644    24743: gap of unknown length
            *    24744    36792: contig of 12049 bp in length
            *    36793    36892: gap of unknown length
            *    36893    59434: contig of 22542 bp in length
            *    59435    59534: gap of unknown length
            *    59535   104072: contig of 44538 bp in length
            *   104073   104172: gap of unknown length
            *   104173   149611: contig of 45439 bp in length
            *   149612   149711: gap of unknown length
            *   149712   256110: contig of 106399 bp in length.
FEATURES             Location/Qualifiers
     source          1. .256110
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-75O6"
     misc_feature    1. .1235
                     /note="assembly_name:Contig16"
     misc_feature    1336. .3639
                     /note="assembly_name:Contig17"
     misc_feature    3740. .5766
                     /note="assembly_name:Contig18"
     misc_feature    5867. .8394
                     /note="assembly_name:Contig19"
     misc_feature    8495. .12007
                     /note="assembly_name:Contig20"
     misc_feature    12108. .17198
                     /note="assembly_name:Contig21"
     misc_feature    17299. .24643
                     /note="assembly_name:Contig22"
     misc_feature    24744. .36792
                     /note="assembly_name:Contig23"
     misc_feature    36893. .59434
                     /note="assembly_name:Contig24"
     misc_feature    59535. .104072
                     /note="assembly_name:Contig25"
     misc_feature    104173. .149611
                     /note="assembly_name:Contig26"
     misc_feature    149712. .256110
                     /note="assembly_name:Contig27"
BASE COUNT    78391 a  49329 c  51220 g  76018 t   1152 others
ORIGIN

  Query Match             17.1%;  Score 170.6;  DB 2;  Length 256110;
  Best Local Similarity   67.0%;  Pred. No. 7.1e-30;
  Matches  242;  Conservative    0;  Mismatches  119;  Indels    0;  Gaps    0;

Qy         85 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 144
              |||   ||  ||  |||  ||   |   || |||||||  ||| ||||||||||
Db     219054 GCTTTGCTTTTTAATGAAAGGCTTGCTTGCTTGCTTGCCTGCTTGCCTGCTTGCCTGCCT 219113

Qy        145 GACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 204
              | | |   || ||| ||||     |   |   || ||| |||| ||| || ||||||||
Db     219114 GCCTGCCTGCTTGCCTGCTTGTCTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTGCTT 219173

Qy        205 AGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTC 264
                |   |   |  |||| ||||||| |   | ||||  ||  |||||| ||| |||
Db     219174 GCTTGCTTGCTTGCTTGCTTGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 219233

Qy        265 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 324
              ||  |||||| ||||||   ||| ||||||||| |||| ||| || ||||||||||
Db     219234 GCCTGCCTGCCTGCTTGTCTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCCT 219293

Qy        325 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 384
              ||| |||||||||| |||| ||| || ||||||||||| ||| || ||||||||||
Db     219294 GCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 219353

Qy        385 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 444
              ||| || |||||||||| |||||  || ||||||    | |   | | || ||  ||  |
Db     219354 GCTTGCTTGCTTGCTTCCTCGCCTGCTTGCTTGGATGGATGGATGGATGGATGGATGGAT 219413

Qy        445 G 445
              |
Db     219414 G 219414


RESULT 55
AC120242/c
```

```
LOCUS       AC120242              201016 bp    DNA     linear   HTG 12-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-379J13, WORKING DRAFT SEQUENCE.
ACCESSION   AC120242
VERSION     AC120242.3  GI:23908271
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 201016)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 201016)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 201016)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 12, 2002 this sequence version replaced gi:21747233.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GVTN
                Center clone name: CH230-379J13
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 186506 bases at least Q40
                Consensus quality: 187889 bases at least Q30
                Consensus quality: 188707 bases at least Q20
                Estimated insert size: 191116; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *         1   201016: contig of 201016 bp in length.
FEATURES             Location/Qualifiers
     source          1. .201016
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-379J13"
     misc_feature    1. .1118
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    7768. .8481
                     /note="clone_boundary
                     clone_end:Sp6
                     site:MboI
                     end_sequence:RXAXC55TV"
     misc_feature    complement(136377. .200482)
                     /note="clone_boundary
                     clone_end:T7
                     site:MboI
                     end_sequence:RXAXC55TJ"
BASE COUNT    51464 a  48878 c  46169 g  42732 t  11773 others
ORIGIN

  Query Match             16.8%;  Score 168;  DB 2;  Length 201016;
  Best Local Similarity   53.1%;  Pred. No. 2.9e-29;
  Matches  357;  Conservative    0;  Mismatches  315;  Indels    0;  Gaps    0;

Qy          64 GGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 123
               ||   ||| |  |||  |   ||||| || | |    | ||   |  ||| || | |
Db      178567 GGCGTGGCCGTGTGCTGGTGTGCTGGTGTGGTGGTGTGGTGGCGTGGTGGCGTGGTGGTG 178508

Qy         124 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCT 183
```

```
Db     178507 TGGTGGCGTGCTGGCGTGGTGGTGTGCTGCTCTGCTGGTGTGGTGGCGTGGTGGTGTGCT 178448

Qy        184 TGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCT 243

Db     178447 GGTGTGCTGGCGTGGTGGTGCGCTGGTGTGCTGGTGTGCTGGTGTGCTGGTGTGGTGGCG 178388

Qy        244 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 303

Db     178387 TGCTGGCGTGGTGGTGTGGTGGCGTGCTGGCGTGGTGGTGTGCTGGTGTGCTCGTGTGCT 178328

Qy        304 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 363

Db     178327 GGTGTGGTGGTGTGGTGGCGTGGTGCTGTGCTGGCGTGGTGGTGTGCTGGTGTGCTGGTG 178268

Qy        364 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGG 423

Db     178267 TCGTGGTGTGGTGGCGTGGTGGTGTGGTAGTGTGCTGGCGTGGTGGCGTGCTGGCGTGGT 178208

Qy        424 GGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTA 483

Db     178207 GGTGTGCTGGCGTGGTGGTGTGCTGGCGTGGTGGTGTGCTGGCGTGGTGGCGTGCTGGCG 178148

Qy        484 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 543

Db     178147 TGGTGGCGTGGTGGCGTGGTGGCGTGCTGGCGTGGTGGTGTGCTGGCGTGCTGGCGTGGT 178088

Qy        544 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 603

Db     178087 GGCGTGCTGGCGTGGTGGTGTGCTGGCGTGGTGGTGTGGTGGCGTGGTGGTGTGCTGGTG 178028

Qy        604 GGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 663

Db     178027 TGGTGGTGTGCTGGTGTGGTGGCATGGTGGTGTGCTGGTGTGGTGGCATGGTGGCGTGGT 177968

Qy        664 GGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGC 723

Db     177967 GGCGTGGTGGCGTGCTGGCGTGGTGGCGTGCTGGCGTGGTGGCGTGCTGGCGTGGTGGCG 177908

Qy        724 CGCTTGCCTGCT 735

Db     177907 TGCTGGCGTGGT 177896


RESULT 56
AC141966
LOCUS       AC141966               197153 bp    DNA     linear   HTG 21-MAR-2003
DEFINITION  Rattus norvegicus clone CH230-372P16, *** SEQUENCING IN PROGRESS
            ***, 73 unordered pieces.
ACCESSION   AC141966
VERSION     AC141966.1  GI:29135423
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 197153)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A.,
            Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
            Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
            Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G.,
            Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
            Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
            Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
            Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
            Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
            Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
            Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
            Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
            Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
            Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
            Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
            Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
            Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
            Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
            Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
            Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
            Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 197153)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
              Center project name: KDFC
              Center clone name: CH230-372P16
            -------------- Summary Statistics
              Sequencing vector: Plasmid;
              Chemistry: Dye-terminator Big Dye: 100% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 165147 bases at least Q40
              Consensus quality: 173510 bases at least Q30
              Consensus quality: 180635 bases at least Q20
              Estimated insert size: 169130; sum-of-contigs estimation
              Quality coverage: 2x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 73 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1049: contig of 1049 bp in length
            *     1050     1149: gap of unknown length
            *     1150     2432: contig of 1283 bp in length
            *     2433     2532: gap of unknown length
```

```
*     2533     3890: contig of 1358 bp in length
*     3891     3990: gap of unknown length
*     3991     5466: contig of 1476 bp in length
*     5467     5566: gap of unknown length
*     5567     7037: contig of 1471 bp in length
*     7038     7137: gap of unknown length
*     7138     8152: contig of 1015 bp in length
*     8153     8252: gap of unknown length
*     8253     9570: contig of 1318 bp in length
*     9571     9670: gap of unknown length
*     9671    11147: contig of 1477 bp in length
*    11148    11247: gap of unknown length
*    11248    12667: contig of 1420 bp in length
*    12668    12767: gap of unknown length
*    12768    13786: contig of 1019 bp in length
*    13787    13886: gap of unknown length
*    13887    14962: contig of 1076 bp in length
*    14963    15062: gap of unknown length
*    15063    16631: contig of 1569 bp in length
*    16632    16731: gap of unknown length
*    16732    17738: contig of 1007 bp in length
*    17739    17838: gap of unknown length
*    17839    19324: contig of 1486 bp in length
*    19325    19424: gap of unknown length
*    19425    21063: contig of 1639 bp in length
*    21064    21163: gap of unknown length
*    21164    22382: contig of 1219 bp in length
*    22383    22482: gap of unknown length
*    22483    23824: contig of 1342 bp in length
*    23825    23924: gap of unknown length
*    23925    25794: contig of 1870 bp in length
*    25795    25894: gap of unknown length
*    25895    27228: contig of 1334 bp in length
*    27229    27328: gap of unknown length
*    27329    28469: contig of 1141 bp in length
*    28470    28569: gap of unknown length
*    28570    29748: contig of 1179 bp in length
*    29749    29848: gap of unknown length
*    29849    31625: contig of 1777 bp in length
*    31626    31725: gap of unknown length
*    31726    32950: contig of 1225 bp in length
*    32951    33050: gap of unknown length
*    33051    34330: contig of 1280 bp in length
*    34331    34430: gap of unknown length
*    34431    36256: contig of 1826 bp in length
*    36257    36356: gap of unknown length
*    36357    37893: contig of 1537 bp in length
*    37894    37993: gap of unknown length
*    37994    39984: contig of 1991 bp in length
*    39985    40084: gap of unknown length
*    40085    41384: contig of 1300 bp in length
*    41385    41484: gap of unknown length
*    41485    43490: contig of 2006 bp in length
*    43491    43590: gap of unknown length
*    43591    45354: contig of 1764 bp in length
*    45355    45454: gap of unknown length
*    45455    46885: contig of 1431 bp in length
*    46886    46985: gap of unknown length
*    46986    48685: contig of 1700 bp in length
*    48686    48785: gap of unknown length
*    48786    50073: contig of 1288 bp in length
*    50074    50173: gap of unknown length
*    50174    51173: contig of 1000 bp in length
*    51174    51273: gap of unknown length
*    51274    52907: contig of 1634 bp in length
*    52908    53007: gap of unknown length
*    53008    55111: contig of 2104 bp in length
*    55112    55211: gap of unknown length
*    55212    57269: contig of 2058 bp in length
*    57270    57369: gap of unknown length
*    57370    58856: contig of 1487 bp in length
*    58857    58956: gap of unknown length
*    58957    61331: contig of 2375 bp in length
*    61332    61431: gap of unknown length
*    61432    63670: contig of 2239 bp in length
*    63671    63770: gap of unknown length
*    63771    66965: contig of 3195 bp in length
*    66966    67065: gap of unknown length
*    67066    68564: contig of 1499 bp in length
*    68565    68664: gap of unknown length
*    68665    70885: contig of 2221 bp in length
*    70886    70985: gap of unknown length
*    70986    73327: contig of 2342 bp in length
*    73328    73427: gap of unknown length
*    73428    77032: contig of 3605 bp in length
*    77033    77132: gap of unknown length
*    77133    79086: contig of 1954 bp in length
*    79087    79186: gap of unknown length
*    79187    80549: contig of 1363 bp in length
*    80550    80649: gap of unknown length
*    80650    82568: contig of 1919 bp in length
*    82569    82668: gap of unknown length
*    82669    85166: contig of 2498 bp in length
*    85167    85266: gap of unknown length
*    85267    87990: contig of 2724 bp in length
*    87991    88090: gap of unknown length
*    88091    90803: contig of 2713 bp in length
*    90804    90903: gap of unknown length
*    90904    94173: contig of 3270 bp in length
*    94174    94273: gap of unknown length

  Query Match            16.7%;  Score 167.4;  DB 2;  Length 197153;
  Best Local Similarity  66.7%;  Pred. No. 4.1e-29;
  Matches  254;  Conservative    0;  Mismatches  126;  Indels    1;  Gaps    1;

Qy         39 GCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGC 98
              |||  |   |||||||  |    |  |   | ||||||||  |||| ||| || ||| |||
Db     111170 GCGCTTTGCTGCTTGTTTGCTTGCTTGCCTGCCTGCCTGCTTGCTTGCTTGCTTGCCTGC 111229

Qy         99 TGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGC 158
                 | | |   | | || |||| | | || ||| |||      |   |   |  |||
Db     111230 CTGCCTGCCTGCTTGTCGGCCTGCTTGTTTGCTTGCCTGCCTGCTTGCTTGCTTGTTTGC 111289

Qy        159 TTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGG 218
              |||||     | | |   |||||||||| ||  ||||||||||     |   ||  |
Db     111290 TTGCTTGCCTGCCTGCCTGCCTGCTTGCTTGCCTGCCTGCTTGCTTGTTTGCTTGTCTGC 111349

Qy        219 CTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 278
              ||||  |||||| ||  |  ||   ||  ||||||||||||||  ||  || |||||||
Db     111350 CTGCCTGCTTGCTTGTTTGCTTGTCTGCCTGCCTGCTTGCTTGCCTGCCTGCTTGCTTGC 111409

Qy        279 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 338
              ||||  ||  ||||||||||||||  ||  ||||||||||||||  ||  |||||| |||
Db     111410 TTGCCTGCCTGCCTGCTTGCTTGCCTGCCTGCCTGCTTGCTTGCCTGCCTGCCTGCCTGC 111469

Qy        339 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 398
               |||  ||| || |||||| ||||| ||| |  ||||||||||||| ||| || || ||||
Db     111470 CTGCCTGCTTGCTTGCTTGTTTGCTTGCTTGTTTGCTTGCTTGCT-GCTTGCTTGTTTGC 111528

Qy        399 TTGCTCGCTGGCCTGCTTGCT 419
              ||||| | | |||||| ||||
Db     111529 TTGCTTGTTTGCCTGCCTGCT 111549


RESULT 57
AC094273/c
LOCUS       AC094273               228645 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-3G15, *** SEQUENCING IN PROGRESS ***,
            2 unordered pieces.
ACCESSION   AC094273
VERSION     AC094273.6  GI:30467504
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
```

```
          Rattus.
REFERENCE 1  (bases 1 to 228645)
  AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
          Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
          Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
          Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
          Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
          Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
          Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
          Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
          Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
          Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
          Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
          Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
          Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
          Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
          Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
          Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
          Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
          Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
          Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
          Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
          Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
          Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
          Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
          Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
          Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
          Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
          Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
          Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
          Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
          Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
          Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
          Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
          Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
          Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
          Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
          Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
          Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
          Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
          Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
          Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
          Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
          Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
          Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
          Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
          Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
          Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
          Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
          Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
          Weinstock,G. and Gibbs,R.A.
  TITLE   Direct Submission
  JOURNAL Unpublished
REFERENCE 2  (bases 1 to 228645)
  AUTHORS Worley,K.C.
  TITLE   Direct Submission
  JOURNAL Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3  (bases 1 to 228645)
  AUTHORS Rat Genome Sequencing Consortium.
  TITLE   Direct Submission
  JOURNAL Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
COMMENT   On May 9, 2003 this sequence version replaced gi:24955409.
          The sequence in this assembly is a combination of BAC based reads
          and whole genome shotgun sequencing reads assembled using Atlas
          (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
          in the feature table below represents a scaffold in the Atlas
          assembly (a 'contig-scaffold'). Within each contig-scaffold,
          individual sequence contigs are ordered and oriented, and separated
          by sized gaps filled with Ns to the estimated size. The sequence
          may extend beyond the ends of the clone and there may be sequence
          contigs within a contig-scaffold that consist entirely of whole
          genome shotgun sequence reads. Both end sequences and whole genome
          shotgun sequence only contigs will be indicated in the feature
          table.
          -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
              Center project name: GAHS
              Center clone name: CH230-3G15
          -------------- Summary Statistics
              Assembly program: Atlas;
              Consensus quality: 217014 bases at least Q40
              Consensus quality: 220400 bases at least Q30
              Consensus quality: 223003 bases at least Q20
              Estimated insert size: 227771; sum-of-contigs estimation
              Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 2 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1   226692: contig of 226692 bp in length
          *  226693   226792: gap of unknown length
          *  226793   228645: contig of 1853 bp in length.
FEATURES             Location/Qualifiers
     source          1. .228645
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-3G15"
     misc_feature    1. .2101
                     /note="wgs_contig"
BASE COUNT    54911 a  52298 c  54612 g  62695 t   4129 others
ORIGIN

  Query Match             16.7%;  Score 167.4;  DB 2;  Length 228645;
  Best Local Similarity   66.7%;  Pred. No. 4e-29;
  Matches  254;  Conservative    0;  Mismatches  126;  Indels    1;  Gaps    1;

Qy         39 GCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGC 98
              |||  |   |||||||  |    |  |   | |||||||| |||| ||| || ||| |||
Db       8080 GCGCTTTGCTGCTTGTTTGCTTGCTTGCCTGCCTGCCTGCTTGCTTGCTTGCTTGCCTGC 8021

Qy         99 TGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGC 158
                 | | |   | | || |||| | | || ||| |||      |   |   | |||
Db       8020 CTGCCTGCCTGCTTGTCGGCCTGCTTGTTTGCTTGCCTGCCTGCTTGCTTGCTTGTTTGC 7961

Qy        159 TTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGG 218
              |||||     | | |   |||||||||| ||  ||||||||||      |   ||  |
Db       7960 TTGCTTGCCTGCCTGCCTGCCTGCTTGCTTGCCTGCCTGCTTGCTTGTTTGCTTGTCTGC 7901

Qy        219 CTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 278
              ||||  |||||| ||  |  ||   ||  |||||||||||||  ||  || ||||||
Db       7900 CTGCCTGCTTGCTTGTTTGCTTGTCTGCCTGCCTGCTTGCTTGCCTGCCTGCTTGCTTGC 7841

Qy        279 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 338
              ||||  ||  ||||||||||||||  ||  ||||||||||||||  ||  |||||| |||
Db       7840 TTGCCTGCCTGCCTGCTTGCTTGCCTGCCTGCCTGCTTGCTTGCCTGCCTGCCTGCCTGC 7781

Qy        339 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 398
              |||  ||| || ||||| ||||| ||| |  ||||||||||| ||| || || ||||
Db       7780 CTGCCTGCTTGCTTGCTTGTTTGCTTGCTTGTTTGCTTGCTTGCT-GCTTGCTTGTTTGC 7722
```

```
Qy        399 TTGCTCGCTGGCCTGCTTGCT 419
              ||||| | | |||||| ||||
Db       7721 TTGCTTGTTTGCCTGCCTGCT 7701


RESULT 58
AC094136/c
LOCUS       AC094136               257675 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-3D20, *** SEQUENCING IN PROGRESS ***,
            3 unordered pieces.
ACCESSION   AC094136
VERSION     AC094136.6  GI:30467923
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 257675)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 257675)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 257675)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 9, 2003 this sequence version replaced gi:24819678.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GABE
                Center clone name: CH230-3D20
            -------------- Summary Statistics
                Assembly program: Atlas;
                Consensus quality: 233178 bases at least Q40
                Consensus quality: 236706 bases at least Q30
                Consensus quality: 239237 bases at least Q20
                Estimated insert size: 245474; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   254381: contig of 254381 bp in length
            *  254382   254481: gap of unknown length
            *  254482   255929: contig of 1448 bp in length
            *  255930   256029: gap of unknown length
            *  256030   257675: contig of 1646 bp in length.
FEATURES             Location/Qualifiers
     source          1. .257675
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-3D20"
     misc_feature    1. .1653
                     /note="wgs_contig"
     misc_feature    6450. .8473
                     /note="wgs_contig"
     misc_feature    223108. .225095
                     /note="wgs_contig"
     misc_feature    complement(246438. .247075)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH308583"
     misc_feature    252628. .254381
                     /note="wgs_end_extension
                     clone_end:T7"
```

```
BASE COUNT    68663 a  49595 c  49881 g  72963 t  16573 others
ORIGIN

  Query Match           16.7%;  Score 167.4;  DB 2;  Length 257675;
  Best Local Similarity   59.6%;  Pred. No. 4e-29;
  Matches  282;  Conservative    0;  Mismatches  191;  Indels    0;  Gaps    0;

Qy          41 GTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTG 100
               |||||  ||| || | | |  || |   |  ||||||| |||  ||  ||||||| |||
Db       88726 GTGTGACTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCT 88667

Qy         101 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 160
                 | | |.|   |||||| |||  ||  |||||| |||    | | | |   |||||| |
Db       88666 GCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGT 88607

Qy         161 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 220
               ||    | | | |   ||||||| |||  ||  |||||| |||    | | | |   | ||
Db       88606 GCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCT 88547

Qy         221 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 280
               ||  || |||  |   |  |||  ||  |||||| ||| |||  ||  || ||| ||| |
Db       88546 GCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGT 88487

Qy         281 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 340
               ||  ||  |||||| ||| |||  ||  || ||| ||| |||  ||  |||||| ||| |
Db       88486 GCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCT 88427

Qy         341 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
               ||  ||  || ||| ||| ||| ||  |||||| ||| |||  ||  || ||| ||| |
Db       88426 GCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGT 88367

Qy         401 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTAC 460
               ||  ||  |||||| |||    | | | |   | |||| ||| |||      |   |
Db       88366 GCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCT 88307

Qy         461 GCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 513
               || ||||||   || ||   |  | |||  |  ||| | |    || | ||
Db       88306 GCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCGACTGCCTCTACAGG 88254


RESULT 59
AC109837/c
LOCUS       AC109837               197891 bp    DNA     linear   HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-271E13, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC109837
VERSION     AC109837.5  GI:24818656
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 197891)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 197891)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 197891)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 9, 2002 this sequence version replaced gi:23101414.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GOGT
                Center clone name: CH230-271E13
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 191506 bases at least Q40
                Consensus quality: 192743 bases at least Q30
                Consensus quality: 193813 bases at least Q20
                Estimated insert size: 195908; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
```

```
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   193483: contig of 193483 bp in length
            *   193484   193583: gap of unknown length
            *   193584   194691: contig of 1108 bp in length
            *   194692   194791: gap of unknown length
            *   194792   196008: contig of 1217 bp in length
            *   196009   196108: gap of unknown length
            *   196109   197891: contig of 1783 bp in length.
FEATURES             Location/Qualifiers
     source          1. .197891
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-271E13"
     misc_feature    1. .1079
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    complement(190998. .191740)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:MboI
                     end_sequence:BZ218549"
     misc_feature    complement(191078. .191745)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:MboI
                     end_sequence:BZ218549"
BASE COUNT    62207 a  36875 c  35461 g  60011 t   3337 others
ORIGIN

  Query Match            16.7%;  Score 166.8;  DB 2;  Length 197891;
  Best Local Similarity  52.6%;  Pred. No. 5.6e-29;
  Matches  363;  Conservative    0;  Mismatches  327;  Indels    0;  Gaps    0;

Qy         44 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 103
              ||  ||  ||    || ||  |   | ||||||||||||  ||  |||||| |||
Db      19486 TGTATGTCTGCCTAACTCCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCC 19427

Qy        104 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 163
              | | |   |||| | |||  |   | |||| |         | | |   |||||| |||
Db      19426 TGCCTGCCTGCCTCCCTGCCTGTATGTCTGCCTAACTCCCTGCCTGCCTGCCTGCGTGCA 19367

Qy        164 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 223
              | | |       || |||  ||  ||  || | | |||      | | ||| | ||||
Db      19366 TACGTGCGTCAGTGCGTGCAAGCCTGCATGCATACATGCATCCCTGCCTGTATGTCTGCA 19307

Qy        224 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 283
              | || |   | ||||  ||      |   || |||  ||  |||| | ||| |||
Db      19306 TAACTCCTTGCCTGCCTGCCTGCCAAAAAGACAGCCTGCCTGCATGCCTACATGCGTGCC 19247

Qy        284 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 343
              |   | |||| |   | |  ||  || ||| ||| |||  ||  |||||| ||| | |
Db      19246 TGTATGTCTGCCTAACTCCCTGCATGCATGCCTGCGTGCATGCATGCCTGCATGCCTACC 19187

Qy        344 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403
              ||  |  ||  ||| |     |   | ||| ||| |||  ||  |||||| ||| |||
Db      19186 TGCCTGTATGTCTGCATAACTCCATACATGCCTGCCTGCCTGCCTGCCTGCCTGCC 19127

Qy        404 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCC 463
              ||  |||||| ||       | |       |||| ||| |||      | | |   |||
Db      19126 TGCCTGCCTGCCTGTATGTCTGCCTAACTCCCTGCCTGCCTGCCTGCCTGCCTGCC 19067

Qy        464 TGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 523
              |||||    || ||   ||  | ||      ||||  |   | ||||  | ||||  |

Db      19066 TGCCTCCCTGCCTGTATGTCTGCCTAACTCCCTGCCTGCCTGCCTGCCTGCCTGCC 19007

Qy        524 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTA 583
              | ||||  | ||||  |      ||||  |  ||   |      ||||  | ||||  |
Db      19006 TGCCTGCCTGCCTGCCTGCCTCCCTGCCTGTATGTCTGCCTAACTCCCTGCCTGCCTGCC 18947

Qy        584 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 643
              | ||||  | ||||  |   | || |  | ||||  |   | ||||  | ||||  |
Db      18946 TGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCC 18887

Qy        644 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCT 703
              | ||      ||||  |   | ||||  | ||||  |   | ||||||||  || |||
Db      18886 TGCCTAACTCCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCC 18827

Qy        704 TGCTTGCTTGCTAGTGGGGCCGCTTGCCTG 733
              ||| ||| |||  |   | |  | || |||
Db      18826 TGCCTGCCTGCCTGCCTGCCTCCCTGGCTG 18797


RESULT 60
AC118905
LOCUS       AC118905               205515 bp    DNA     linear   HTG 12-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-255D17, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC118905
VERSION     AC118905.7  GI:23908079
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 205515)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
```

```
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 205515)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 205515)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 12, 2002 this sequence version replaced gi:21902607.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GWJB
                Center clone name: CH230-255D17
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 184541 bases at least Q40
                Consensus quality: 186580 bases at least Q30
                Consensus quality: 188187 bases at least Q20
                Estimated insert size: 190219; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    22234: contig of 22234 bp in length
            *    22235    22334: gap of unknown length
            *    22335   154198: contig of 131864 bp in length
            *   154199   154298: gap of unknown length
            *   154299   179162: contig of 24864 bp in length
            *   179163   179262: gap of unknown length
            *   179263   205515: contig of 26253 bp in length.
FEATURES            Location/Qualifiers
     source         1. .205515
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-255D17"
     misc_feature   1. .1336
                    /note="wgs_end_extension
                    clone_end:T7"
     misc_feature   complement(938. .1516)
                    /note="clone_boundary
                    clone_end:T7
                    site:MboI
                    end_sequence:RXACK21TJ"
     misc_feature   51900. .52907
                    /note="wgs_contig"
BASE COUNT    56529 a  39946 c  39535 g  53910 t  15595 others
ORIGIN

  Query Match           16.6%;  Score 166.4;  DB 2;  Length 205515;
  Best Local Similarity   64.6%;  Pred. No. 7e-29;
  Matches  248;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy        132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191
              ||  |||      | | |   ||||||||||      | | ||  ||||| |||| |||
Db     121923 TGTCTGCCTGTTTGCCTGCCTGCCTGCTTGCCTGTTTGCCTGTTTGCCTGTTTGCCTGCT 121982

Qy        192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251
              || |  |   ||   |     | |     || | ||||  |   | |||||| ||  ||
Db     121983 TGCGTCTTGCTTGCTTGCCTGCTTGCTTGCCTTGTTTGCCTGCTTGCCTGCTTGCCTGCT 122042

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
              ||||||||||   ||| ||  |||||| |||| ||| || ||| |||||||| ||  |||
Db     122043 TGCTTGCTTGTCTGCTTGCTCGCTTGCCTGCTTGCTCGCTTGCCTGCTTGCTTGCCTGCC 122102

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
              ||| ||||||||||| |||||||||||||| ||| || ||| |||||||  ||| |||
Db     122103 TGCCTGCTTGCTCGCTTGCCTGCTTGCTTGCTTGCTTGCCTGCTTGCCTGCTTGCC 122162

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTA 431
              |||||||| |||| ||| || ||| |||||||  ||  ||||||||||||     | | |
Db     122163 TGCTTGCCTGCTTGCTCGCTTGCCTGCTTGCCTGCCTGCCTGCTTGCTCGCTCGCCTGCT 122222

Qy        432 GGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCT 491
              | ||||||||||||||     |   |   || |||||| | ||||| |     |  ||||
Db     122223 CGCCTGCTTGCTTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCTTGCTCACTCGCTTGCT 122282

Qy        492 GGGCTGCTAGTAGGGCTGCTGGGC 515
              |   |||  |   |  || ||||
Db     122283 TGCTCGCTCATTTTGAAGCAGGGC 122306


RESULT 61
AC109709
LOCUS       AC109709              272652 bp    DNA     linear   HTG 09-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-228K2, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
ACCESSION   AC109709
VERSION     AC109709.4  GI:23195460
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 272652)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
```

```
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 272652)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 272652)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 19, 2002 this sequence version replaced gi:21738042.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GQEA
                Center clone name: CH230-228K2
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 233404 bases at least Q40
                Consensus quality: 236957 bases at least Q30
                Consensus quality: 239428 bases at least Q20
                Estimated insert size: 252351; sum-of-contigs estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   170363: contig of 170363 bp in length
            *   170364   170463: gap of unknown length
            *   170464   271425: contig of 100962 bp in length
            *   271426   271525: gap of unknown length
            *   271526   272652: contig of 1127 bp in length.
FEATURES             Location/Qualifiers
     source          1. .272652
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-228K2"
     misc_feature    complement(95812. .96229)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:RWBNV61TJ"
     misc_feature    152380. .153698
                     /note="wgs_contig"
     misc_feature    complement(168487. .169398)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:RWBNV61TV"
     misc_feature    170464. .172041
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    266872. .268448
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    268995. .271425
                     /note="wgs_end_extension
                     clone_end:Sp6"
BASE COUNT    71259 a  49098 c  50266 g  71245 t  30784 others
ORIGIN

  Query Match            16.6%;  Score 166.4;  DB 2;  Length 272652;
  Best Local Similarity  64.6%;  Pred. No. 6.8e-29;
  Matches  248;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy        132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191
              ||  |||      | | |   ||||||||||      | | ||  ||||| ||||  |||
Db     140422 TGTCTGCCTGTTTGCCTGCCTGCCTGCTTGCCTGTTTGCCTGTTTGCCTGTTTGCCTGCT 140481

Qy        192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251
               || |  |   ||   |     | |     || | ||||  |   | |||||  ||  ||
Db     140482 TGCGTCTTGCTTGCTTGCCTGCTTGCTTGCCTTGTTTGCCTGCTTGCCTGCTTGCCTGCT 140541

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
              ||||||||||   ||| ||  ||||||  |||| ||| || ||| ||||||||  ||  |||
Db     140542 TGCTTGCTTGTCTGCTTGCTCGCTTGCCTGCTTGCTCGCTTGCCTGCTTGCTTGCCTGCC 140601

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
              ||| ||||||||||||| ||||||||||||||| ||| || ||| |||||||  ||| |||
Db     140602 TGCCTGCTTGCTCGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCCTGCTTGCCTGCTTGCC 140661
```

```
Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTA 431
              ||||||| |||| ||| || ||| ||||||| || |||||||||||     | | |
Db     140662 TGCTTGCCTGCTTGCTCGCTTGCCTGCTTGCCTGCCTGCCTGCTTGCTCGCTCGCCTGCT 140721

Qy        432 GGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCT 491
              | |||||||||||||     |   |   || |||||| | ||||| |     |  ||||
Db     140722 CGCCTGCTTGCTTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCTTGCTCACTCGCTTGCT 140781

Qy        492 GGGCTGCTAGTAGGGCTGCTGGGC 515
              |   |||  |   |  || ||||
Db     140782 TGCTCGCTCATTTTGAAGCAGGGC 140805
```

```
RESULT 62
AC103472
LOCUS       AC103472               234474 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-208G21, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC103472
VERSION     AC103472.7  GI:30520476
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 234474)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R.,.Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 234474)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 234474)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:25092334.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GHRX
                Center clone name: CH230-208G21
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 223543 bases at least Q40
                Consensus quality: 225303 bases at least Q30
                Consensus quality: 226470 bases at least Q20
                Estimated insert size: 235136; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   233348: contig of 233348 bp in length
            *   233349   233448: gap of unknown length
            *   233449   234474: contig of 1026 bp in length.
FEATURES             Location/Qualifiers
     source          1. .234474
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-208G21"
     misc_feature    1. .2197
                     /note="wgs_contig"
BASE COUNT    62773 a  47967 c  50362 g  65964 t   7408 others
ORIGIN
```

```
  Query Match             16.5%;  Score 165.2;  DB 2;  Length 234474;
  Best Local Similarity   52.6%;  Pred. No. 1.3e-28;
```

```
  Matches  359;  Conservative    0;  Mismatches  323;  Indels    0;  Gaps    0;

Qy         65 GTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 124
              ||  |  |||||||  ||   ||  ||||  |  ||      | | | | | || ||| |||
Db      50013 GTGTGTGTGCCTGCGTGTCTGCGCGCCTACGTGTCAGCGTGTCTGCATGCTTGCATGCCT 50072

Qy        125 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 184
              ||  | |||    |  | | | | |   ||||||  ||        |   |   ||||||  |
Db      50073 GCGTGTCTGTGCACCTATGTGCCTGCGTGCCTGCCTGTGTGCCTGCGTGCCTGCCTGCGT 50132

Qy        185 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTC 244
              ||  ||  || ||| |||    | | | |   |  |||  || |||  |  . | ||||
Db      50133 GCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGT 50192

Qy        245 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 304
              ||  |||||| ||| ||  ||  |||||| ||| ||  ||  || ||| ||| |||
Db      50193 GCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGT 50252

Qy        305 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 364
              ||  |||||| ||| ||   |  |||| ||   |  ||   || ||| ||| ||
Db      50253 GCCTGCCTGCGTGCCTGTCAGTGCGTCTGCGTGTCAGTGCGTCTGCGTGCCTGCGTGTCT 50312

Qy        365 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 424
              ||  |||||| ||| |||  | | |||| | |||||||  |   || ||  |||    |
Db      50313 GCGTGCCTGCGTGCCTGCGTGTTTGCCTACGTGCTTGCGTGTCTGCGTGTGTGCCTGCGT 50372

Qy        425 GACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAG 484
              | |     | || | ||| |         |   |  || ||||||  |  |  ||
Db      50373 GCCTACGTGTCTACGTGCCTATGTGTCAGCGTGCCTGCGTGCCTGCGTGTCAGCGTGTCA 50432

Qy        485 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 544
              |   |  |  ||| |  | ||||  | ||||  ||  |  |||  |  ||   |
Db      50433 GTGCGTCTGCGTGCCTGCGTGTCTGCGTGCCTGCGTGTCTGTATGCCTGCGTGTCTGCGT 50492

Qy        545 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 604
              | ||||  | ||||  |      |   |  |||      | ||||  | ||||  ||
Db      50493 GCCTGCGTGTCTGCGTGCCTCCGTGTCTGCGTGCCTCCGTGTCTGCGTGCCTGCGTGTCT 50552

Qy        605 GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 664
              |  |  |  ||   ||| | ||||  | ||||  |   |  ||   |  |||  |
Db      50553 GTATGCCTGCGTGTCTGTATGCCTGCGTGTCTGCGTGCCTGCGTGTCTGCGTGCCTGCGT 50612

Qy        665 GGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCC 724
              | ||||  | || |  ||  |  | ||||   |  ||| ||| ||| ||   |  | |
Db      50613 GTCTGCGTGCCTCCGTGTCTGCGTACCTGTGTGTCTGCGTGCCTGCGTGTCTGCATGTCT 50672

Qy        725 GCTTGCCTGCTACTAGGGCTGC 746
              || |||||||   |  |  |||
Db      50673 GCGTGCCTGCATGTGTGCTTGC 50694


RESULT 63
AC112613/c
LOCUS       AC112613               235837 bp    DNA     linear   HTG 13-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-101I20, WORKING DRAFT SEQUENCE, 6
            unordered pieces.
ACCESSION   AC112613
VERSION     AC112613.5  GI:24941847
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 235837)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 235837)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 235837)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 13, 2002 this sequence version replaced gi:23266005.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
```

```
             ---------- Project Information
                 Center project name: GRSW
                 Center clone name: CH230-101I20
             ---------- Summary Statistics
                 Assembly program: Phrap; version 0.990329
                 Consensus quality: 223584 bases at least Q40
                 Consensus quality: 225354 bases at least Q30
                 Consensus quality: 226739 bases at least Q20
                 Estimated insert size: 235335; sum-of-contigs estimation
                 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
             ----------
             * NOTE: Estimated insert size may differ from sequence length
             *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 6 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *       1   184317: contig of 184317 bp in length
             *  184318   184417: gap of unknown length
             *  184418   186305: contig of 1888 bp in length
             *  186306   186405: gap of unknown length
             *  186406   188834: contig of 2429 bp in length
             *  188835   188934: gap of unknown length
             *  188935   192034: contig of 3100 bp in length
             *  192035   192134: gap of unknown length
             *  192135   214096: contig of 21962 bp in length
             *  214097   214196: gap of unknown length
             *  214197   235837: contig of 21641 bp in length.
FEATURES             Location/Qualifiers
     source          1. .235837
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-101I20"
BASE COUNT    69431 a  48015 c  45986 g  63690 t   8715 others
ORIGIN

  Query Match             16.5%;  Score 165.2;  DB 2;  Length 235837;
  Best Local Similarity   52.6%;  Pred. No. 1.3e-28;
  Matches  359;  Conservative    0;  Mismatches  323;  Indels    0;  Gaps    0;

Qy          65 GTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 124
               ||  |  ||||||| ||   ||  |||| | ||     | | | | | || ||| |||
Db      164965 GTGTGTGTGCCTGCGTGTCTGCGCGCCTACGTGTCAGCGTGTCTGCATGCTTGCATGCCT 164906

Qy         125 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 184
               ||  | |||    |  | | | | |   |||||| ||       |   |   |||||| |
Db      164905 GCGTGTCTGTGCACCTATGTGCCTGCGTGCCTGCCTGTGTGCCTGCGTGCCTGCCTGCGT 164846

Qy         185 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTC 244
               ||  ||  || ||| |||    | | | |   |  |||  || |||  |   | ||||
Db      164845 GCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGT 164786

Qy         245 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 304
               ||  |||||| ||| |||  ||  |||||| ||| |||  ||  || ||| ||| |||
Db      164785 GCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGT 164726

Qy         305 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 364
               ||  |||||| ||| ||   |  | |||| ||   |  ||   || ||| ||| ||
Db      164725 GCCTGCCTGCGTGCCTGTCAGTGCGTCTGCGTGTCAGTGCGTCTGCGTGCCTGCGTGTCT 164666

Qy         365 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 424
               ||  |||||| ||| |||  | | |||| | |||||||  |   || ||  |||    |
Db      164665 GCGTGCCTGCGTGCCTGCGTGTTTGCCTACGTGCTTGCGTGTCTGCGTGTGTGCCTGCGT 164606

Qy         425 GACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAG 484
               | |     | || | ||| |        |   |   || ||||||   |   |   ||
Db      164605 GCCTACGTGTCTACGTGCCTATGTGTCAGCGTGCCTGCGTGCCTGCGTGTCAGCGTGTCA 164546

Qy         485 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 544
               |   |   |  |||  |   | ||||  | ||||  || .|  |||  |  ||   |
Db      164545 GTGCGTCTGCGTGCCTGCGTGTCTGCGTGCCTGCGTGTCTGTATGCCTGCGTGTCTGCGT 164486

Qy         545 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 604
               | ||||  | ||||  |      |   |  |||     .| ||||  | ||||  ||
Db      164485 GCCTGCGTGTCTGCGTGCCTCCGTGTCTGCGTGCCTCCGTGTCTGCGTGCCTGCGTGTCT 164426

Qy         605 GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 664
               |  | |  |  ||   ||| | |||| | ||||  |   |  ||   |  |||  |
Db      164425 GTATGCCTGCGTGTCTGTATGCCTGCGTGTCTGCGTGCCTGCGTGTCTGCGTGCCTGCGT 164366

Qy         665 GGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCC 724
               | ||||  | || |  ||  |  | ||||   |  ||| ||| ||| ||  |   | |
Db      164365 GTCTGCGTGCCTCCGTGTCTGCGTACCTGTGTGTCTGCGTGCCTGCGTGTCTGCATGTCT 164306

Qy         725 GCTTGCCTGCTACTAGGGCTGC 746
               || |||||||   |  |  |||
Db      164305 GCGTGCCTGCATGTGTGCTTGC 164284


RESULT 64
AC106350/c
LOCUS       AC106350               256486 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-94P16, WORKING DRAFT SEQUENCE, 6
            unordered pieces.
ACCESSION   AC106350
VERSION     AC106350.4  GI:30581308
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 256486)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angulano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
```

Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 256486)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 256486)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 13, 2003 this sequence version replaced gi:23268877.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
-------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
Center project name: GKSS
Center clone name: CH230-94P16
-------------- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 236501 bases at least Q40
Consensus quality: 239568 bases at least Q30
Consensus quality: 241652 bases at least Q20
Estimated insert size: 246612; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 11137: contig of 11137 bp in length
* 11138 11237: gap of unknown length
* 11238 249465: contig of 238228 bp in length
* 249466 249565: gap of unknown length
* 249566 250639: contig of 1074 bp in length
* 250640 250739: gap of unknown length
* 250740 251829: contig of 1090 bp in length
* 251830 251929: gap of unknown length
* 251930 253165: contig of 1236 bp in length
* 253166 253265: gap of unknown length
* 253266 256486: contig of 3221 bp in length.

FEATURES Location/Qualifiers
source 1. .256486
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-94P16"
misc_feature 1. .1811
/note="wgs_contig"
misc_feature 11238. .12909
/note="wgs_contig"
misc_feature 243988. .246089
/note="wgs_contig"
misc_feature 247863. .249465
/note="wgs_contig"
BASE COUNT 65807 a 51132 c 53326 g 72818 t 13403 others
ORIGIN

Query Match 16.5%; Score 165.2; DB 2; Length 256486;
Best Local Similarity 52.5%; Pred. No. 1.3e-28;
Matches 362; Conservative 0; Mismatches 328; Indels 0; Gaps 0;

```
Qy        69 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 128
             || ||  ||  || | | |    || ||| |  |  |   |   |  || ||| | | |
Db     83980 GGTTGGTTGGTTGGTTGGTTTGATGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 83921

Qy       129 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 188
             |  || ||      |  |   |   |  || ||| |  |  |   |   |  || |||
Db     83920 GGTTGGTTTGATGGTTGGATGGTTGGTTGGTTGGTTGGATGGTTGGTTGGTTGGTTGGAT 83861

Qy       189 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 248
             |   | ||  || | |  |  |   | | || || | | ||| | |   || || | | |
Db     83860 GGATGGTTGGATGGTTGGTTGGTTGGATGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTT 83801

Qy       249 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 308
              .|||||| ||| | | | |  || ||| ||| | | | |  ||  ||  || | |
Db     83800 TGATGCTTGGTTGGTTGGTTGGGTGGTTGGTTGGTTGTTTGGTTGGATGGATGGTTGGAT 83741

Qy       309 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 368
             |  || ||| ||| | ||| || |||||||||||| ||| || ||  |  ||| | | |
Db     83740 GGTTGGTTGGTTGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGGATTGTTGGTTGGTT 83681

Qy       369 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 428
             |  || ||| ||| | | | |  || ||| ||| | |   |  || ||| |  |  |
Db     83680 GGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGATGGATGGTTGGTTGGTTGGTT 83621

Qy       429 GTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCT 488
             | | || || |||  || |  |  |   | | |  ||  || | | |||   |  || |
Db     83620 GGATGGATGGTTGGATGGTTGGTTGGTTGGATGGTTGGTTGGTTGGTTGGCTGGTTGGTT 83561

Qy       489 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 548
             | |  | |||| |    || || | || || | |   || |||| || ||   | | || |
Db     83560 GGTTTGATGCTTGGTTGGTTGGTTGGGTGGTTGGTTGGTTGCTTGGTTGGATGGATGGTT 83501

Qy       549 GCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 608
             ||  || || | |   |  | || |  |||| |   |  |||| |  |||| |   || |
Db     83500 GCATGGTTGGTTGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGGATGGAT 83441

Qy       609 CCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 668
               | || || | |   || || | || || | |   || || | || ||   |  || |
Db     83440 TGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGATGGATGGTTGGTT 83381

Qy       669 GCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTT 728
             | | || || | | | || ||  || | | ||| |||  || ||| | |   ||  | |
Db     83380 GGTTGGATGGTTGGATGGTTGGTTGGTTGGTTGGTTGGATGGTTGGTTGGTTGGTTGGAT 83321

Qy       729 GCCTGCTACTAGGGCTGCTGTGCAGCTGGG 758
             |  ||       || || |  | | | ||
Db     83320 GGTTGGCTGGTTGGATGATTGGATGGTTGG 83291
```

RESULT 65

```
AC122009
LOCUS       AC122009              152054 bp    DNA     linear   ROD 16-JUL-2002
DEFINITION  Mus musculus clone RP24-347B22, complete sequence.
ACCESSION   AC122009
VERSION     AC122009.2  GI:21844644
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 152054)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 152054)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 152054)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 152054)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Jul 16, 2002 this sequence version replaced gi:21040133.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information -----------------
            Center project name: M_BB0347B22
            ------------------------------------------------------
FEATURES             Location/Qualifiers
     source          1. .152054
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-347B22"
BASE COUNT    44946 a  33300 c  31206 g  42602 t
ORIGIN

  Query Match           16.5%;  Score 165;  DB 10;  Length 152054;
  Best Local Similarity   64.0%;  Pred. No. 1.5e-28;
  Matches  249;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy          63 GGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGC 122
               || ||   |||| ||| |||| ||| || ||||||||     |   |   || |||||||
Db      126484 GGTTAACTCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 126543

Qy         123 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGC 182
               | ||| || ||| |   |  |   ||     | |||||||      | | |   || |||
Db      126544 TTGCTTGCTTGCCTATGGGTGTTCCATGGTTCTTGCTTGCCTGCTTGCCTGCTTGCTTGC 126603

Qy         183 TTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGC 242
               |||||  ||| || ||||||||     |   |   |  |||| ||||||| |       |
Db      126604 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTATGGGT 126663

Qy         243 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 302
                 |     | |||||||||  ||| ||||||||||||||| ||| || ||||||||||
Db      126664 GTTCCATGGTTCTTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGC 126723

Qy         303 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 362
               | ||| || ||||||||||| ||| || ||||||||||     ||  | |    | |
Db      126724 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTATGGGTGTTCCATGGTTC 126783

Qy         363 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAG 422
               | ||| |||||||||| |||| ||| || |||||||||||| ||| || ||||||||
Db      126784 TTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 126843

Qy         423 GGGACAGTAGGGCTGCTTGCTTGCTAAGG 451
                 | | || ||  | |     | | || |
Db      126844 TTGCCTGTGGGTGTTCCATGGTTCCAATG 126872


RESULT 66
AC113245
LOCUS       AC113245              210964 bp    DNA     linear   HTG 06-MAR-2003
DEFINITION  Mus musculus clone RP23-175D5, WORKING DRAFT SEQUENCE, 16 unordered
            pieces.
ACCESSION   AC113245
VERSION     AC113245.5  GI:28867081
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 210964)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-175D5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 210964)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 210964)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2003) Whitehead Institute/MIT Center for Genome
```

```
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 6, 2003 this sequence version replaced gi:28629993.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L23486
                Center clone name: 175_D_5
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 206691 bases at least Q40
                Consensus quality: 208198 bases at least Q30
                Consensus quality: 208925 bases at least Q20
                Insert size: 220000; agarose-fp
                Insert size: 209464; sum-of-contigs
                Quality coverage: 8.3 in Q20 bases; agarose-fp
                Quality coverage: 8.7 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      622: contig of 622 bp in length
            *      623      722: gap of 100 bp
            *      723     1420: contig of 698 bp in length
            *     1421     1520: gap of 100 bp
            *     1521     2183: contig of 663 bp in length
            *     2184     2283: gap of 100 bp
            *     2284     2935: contig of 652 bp in length
            *     2936     3035: gap of 100 bp
            *     3036     4527: contig of 1492 bp in length
            *     4528     4627: gap of 100 bp
            *     4628     8428: contig of 3801 bp in length
            *     8429     8528: gap of 100 bp
            *     8529    12863: contig of 4335 bp in length
            *    12864    12963: gap of 100 bp
            *    12964    20933: contig of 7970 bp in length
            *    20934    21033: gap of 100 bp
            *    21034    34850: contig of 13817 bp in length
            *    34851    34950: gap of 100 bp
            *    34951    49577: contig of 14627 bp in length
            *    49578    49677: gap of 100 bp
            *    49678    65321: contig of 15644 bp in length
            *    65322    65421: gap of 100 bp
            *    65422    84391: contig of 18970 bp in length
            *    84392    84491: gap of 100 bp
            *    84492   111661: contig of 27170 bp in length
            *   111662   111761: gap of 100 bp
            *   111762   139035: contig of 27274 bp in length
            *   139036   139135: gap of 100 bp
            *   139136   178376: contig of 39241 bp in length
            *   178377   178476: gap of 100 bp
            *   178477   210964: contig of 32488 bp in length.
FEATURES              Location/Qualifiers
     source           1. .210964
                      /organism="Mus musculus"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:10090"
                      /clone="RP23-175D5"
                      /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature     1. .622
                      /note="assembly_fragment"
     misc_feature     723. .1420
                      /note="assembly_fragment"
     misc_feature     1521. .2183
                      /note="assembly_fragment"
     misc_feature     2284. .2935
                      /note="assembly_fragment"
     misc_feature     3036. .4527
                      /note="assembly_fragment"
     misc_feature     4628. .8428
                      /note="assembly_fragment"
     misc_feature     8529. .12863
                      /note="assembly_fragment"
     misc_feature     12964. .20933
                      /note="assembly_fragment"
     misc_feature     21034. .34850
                      /note="assembly_fragment"
     misc_feature     34951. .49577
                      /note="assembly_fragment"
     misc_feature     49678. .65321
                      /note="assembly_fragment"
     misc_feature     65422. .84391
                      /note="assembly_fragment"
     misc_feature     84492. .111661
                      /note="assembly_fragment"
     misc_feature     111762. .139035
                      /note="assembly_fragment"
     misc_feature     139136. .178376
                      /note="assembly_fragment"
     misc_feature     178477. .210964
                      /note="assembly_fragment
                      clone_end:T7
                      vector_side:right"
BASE COUNT    60649 a  44731 c  44401 g  59683 t   1500 others
ORIGIN

  Query Match             16.5%;  Score 164.8;  DB 2;  Length 210964;
  Best Local Similarity   48.2%;  Pred. No. 1.6e-28;
  Matches  295;  Conservative    0;  Mismatches  317;  Indels    0;  Gaps    0;

Qy        216 GGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 275
              ||| |||||| ||||| | |     | | |||||  |||| | ||||| | | ||  | |
Db       7923 GGGTTGCTGGGTTGCTGGGATACTGGGTTGCTGGGTTGCTGGGTTGCTGGGTTGCTGGGT 7982

Qy        276 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 335
              |||| | | ||| |  |||| |  | || | ||||  | | ||| | | |||||  | ||
Db       7983 TGCTGGGTTGCTAGGTTGCTGGGATACTGGGTGGCTGGGTGGCTGGGTTGCTGGGATACT 8042

Qy        336 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 395
              | ||||| | ||||  | | ||| | | |||||   ||| | | ||| | ||||  | |
Db       8043 GGGTTGCTGGGTGGCTGGGTGGCTGGGTGGCTGGGTGGCTGGGTGGCTGGATGGCTGGGT 8102

Qy        396 TGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGAC 455
              |||| | | |||||  |||| | | |  ||   |  ||| |||| | |||||  |
Db       8103 TGCTGGGTTGCTGGGTTGCTGGGTTACTGGGTTGCTGGGTTGCTAGGTTGCTGGGATACT 8162

Qy        456 GGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGC 515
              ||   ||  |  || ||| ||| | ||  |   | |     | | |  ||| |||||||
Db       8163 GGGTGGCTGGGTTGCTGGGTTGCTGGTTTGCTGGGTTACTGGGTTGCTGGGTTGCTGGGT 8222

Qy        516 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGC 575
              ||||||   |   |       | | |  ||| ||||||| |||| |         | |
Db       8223 TGCTAGGTTGCTGGGATACTGGGTGGCTGGGTTGCTGGGTTGCTGGGTTACTGGGTTGCT 8282

Qy        576 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGC 635
              | | |  ||| | ||||| |||| |   |     | |   | | |  ||| | |||||
Db       8283 GGGTTGCTGGGTTACTGGGTTGCTGGGTTGCTGGGTTGCTAGGTTGCTGGGATACTGGGT 8342

Qy        636 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCT 695
              |||  |  ||   | |||   | | |  ||| | ||||| |||| |  ||   | | |||
Db       8343 GGCTGGGTGGCTGGGTGGCTGGGTTGCTGGGATACTGGGTTGCTGGGTGGCTGGGCGGCT 8402

Qy        696 GGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCT 755
              || | ||| | | ||| | | |  ||
```

```
Db      8403 GGATGGCTGGATGGCTGGGTGGCTGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 8462

Qy       756 GGGAGAACAGAGTAGGGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAG 815

Db      8463 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 8522

Qy       816 GACACGGGACCC 827
                 | | |
Db      8523 NNNNNNGCTCAC 8534


RESULT 67
AC128183
LOCUS       AC128183              196519 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-381B8, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC128183
VERSION     AC128183.3  GI:25081459
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 196519)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196519)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 196519)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:22855982.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GZNP
                Center clone name: CH230-381B8
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 150327 bases at least Q40
                Consensus quality: 152783 bases at least Q30
                Consensus quality: 154221 bases at least Q20
                Estimated insert size: 156901; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   192049: contig of 192049 bp in length
            *  192050   192149: gap of unknown length
            *  192150   196519: contig of 4370 bp in length.
FEATURES             Location/Qualifiers
     source          1. .196519
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-381B8"
     misc_feature    1. .1529
                     /note="wgs_contig"
     misc_feature    1996. .3760
                     /note="wgs_contig"
     misc_feature    6562. .8382
                     /note="wgs_contig"
     misc_feature    188709. .192049
                     /note="wgs_contig"
BASE COUNT    45700 a  35243 c  33203 g  41414 t  40959 others
ORIGIN
```

```
  Query Match             16.4%;  Score 164;  DB 2;  Length 196519;
  Best Local Similarity   55.4%;  Pred. No. 2.6e-28;
  Matches  317;  Conservative    0;  Mismatches  255;  Indels    0;  Gaps    0;

Qy        173 GTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTA 232
              ||  |  || ||| | | | | ||| ||| |  |  | | |   || || | | |||
Db      89840 GTTTGTTTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGCT 89899

Qy        233 GTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 292
              |   || || | | | |  || ||| ||| | |   |  ||  || |||     | | ||
Db      89900 GGTTGGTTGGTTGGTTGGTTGGTTGATTGGTTGTCTGGTTGTCTGGTTGTCTAGTTGTCT 89959

Qy        293 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 352
              | ||| |||   | | | ||| |||  || | |   |  ||  || |||   | | | ||
Db      89960 GGTTGGTTGGCTGGTTGTCTGGTTGGCTGGTTGTCTGGTTGGCTGGTTGTCTGGTTGGCT 90019

Qy        353 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 412
              | ||| ||| | | | | ||| ||| ||| |   | | | | ||| ||| | | | | ||
Db      90020 GGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTTTGTTGGCCGGTTGGTTGGTTGGTTGGCT 90079

Qy        413 GCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATG 472
              | ||| |  |  | | |   || || ||| |||    |  |   |   |  || ||| |
Db      90080 GGTTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGCTGGTT 90139

Qy        473 GCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 532
              | ||| | |   |||||   || || | |   || ||   || |||| |   || || |
Db      90140 GGTTGGTTGGTTGGCTGGCTGGTTGGTTGGTTGGTTGGCTGGTTGCTTGGCTGGTTGGTT 90199

Qy        533 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTG 592
              || |  | |   || || | || || | |   || |  | ||||| | |   ||||| |
Db      90200 GGTTTGTTGGCTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGCTGGTT 90259

Qy        593 GGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 652
              |||||  | |   || |  | | ||| | |   || || | ||||| | |   ||||| |
Db      90260 GGCTGGTTGGTTGGTTGGTTGACTGGTTGGTTGGTTGGTTGGCTGGTTGCTTGGCTGGTT 90319

Qy        653 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTT 712
              || || | |   |||||   || || | |   ||||| |||   | ||| ||| ||| ||
Db      90320 GGTTGGTTGGTTGGCTGGCTGGTTGGTTGGTTGGCTGGCTGGCTGGTTGGTTGGTTGGTT 90379

Qy        713 GCTAGTGGGGCCGCTTGCCTGCTACTAGGGCT 744
              | | |   ||  | |||  || |  |  ||||
Db      90380 GGTTGGTTGGTTGGTTGGTTGGTTTTGAGGCT 90411


RESULT 68
AC095211
LOCUS       AC095211               252801 bp    DNA     linear   HTG 03-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-9E3, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC095211
VERSION     AC095211.4  GI:22773101
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 252801)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 252801)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 252801)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 10, 2002 this sequence version replaced gi:21722802.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequening reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/).  As a result, the
            sequence may extend beyond the ends of the clone and there may be
            contigs that consist entirely of whole genome shotgun sequence
            reads.  Both end sequences and whole genome shotgun sequence only
            contigs will be indicated in the feature table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GDHZ
                Center clone name: CH230-9E3
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 222989 bases at least Q40
                Consensus quality: 225848 bases at least Q30
                Consensus quality: 227780 bases at least Q20
                Estimated insert size: 241532; sum-of-contigs estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
```

```
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   252801: contig of 252801 bp in length.
FEATURES             Location/Qualifiers
     source          1. .252801
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-9E3"
     misc_feature    1. .1620
                     /note="wgs_contig"
     misc_feature    251753. .252801
                     /note="wgs_contig"
BASE COUNT    63531 a  49651 c  49432 g  66454 t  23733 others
ORIGIN

  Query Match             16.4%;  Score 163.8;  DB 2;  Length 252801;
  Best Local Similarity   54.2%;  Pred. No. 2.8e-28;
  Matches  353;  Conservative    0;  Mismatches  297;  Indels    1;  Gaps    1;

Qy         69 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 128
              ||||| ||  ||   | | |  || |||    |  |   |   |  || ||| | | |
Db     185218 GGCTGGTTGGTTGGCTGGTTGGTTGGTTGGCTGGCTGGTTGGTTGGTTGGTTGGTTGGTT 185277

Qy         129 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 188
               |  || ||| |  |  |   |  || ||| ||| |  |       |   |  || ||| |
Db     185278 GGTTGGTTGGTTGGTTGGTTGGTGGGCTGGTTGGTTGGTCAGTCGGTTGGTTGGTTGGTG 185337

Qy         189 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 248
               |   |  || ||| |  |  |   |   || || | | ||| | |   || || | | |
Db     185338 GGCTGGTTGGTTGGTTGGTTGGTGGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 185397

Qy         249 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 308
               | ||| ||| ||| ||  | |  || ||| |||   | | |  ||  || ||| | |
Db     185398 GGCTGGTTGGTTGGTCAGTCGGTTGGTTGGTTGGCTGGTTGGTTGACTGGTTGGTTGGCT 185457

Qy         309 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 368
               |  |||||| |||   | | || || | |  || | | | | ||| ||| ||| | | ||
Db     185458 GGTTGCTTGGTTGGCTGGTTGCTTGGTAGGCTGATTGGTTGGCTGGTTGGTTGGTTGGTG 185517

Qy         369 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 428
               | ||| ||| |||   | | || || ||| |||||  | ||| ||| |||||  | || |
Db     185518 GGCTGATTGGTTGGCTGGTTGCTTGGTTGGTTGCTTGGTGGGCTGGTTGCTTGGTGGGCT 185577

Qy         429 GTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCT 488
                  || ||  || |||||  |  | | |   | |||   | |||   | |  | .||
Db     185578 TGTTGGTTGGCTGGTTGCTTGGTTGGCTGGTGGGCTGTTGGTTGGTGGGCTGATTGGTTG 185637

Qy         489 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 548
               |||||   ||| || ||  || | || || | |   || || | ||||| | |   ||
Db     185638 GCTGGTTGGCTGGTTGGT-TGGTTGGTTGGTGGGCTGGTTGTTTGGCTGGTTGGTTGGTG 185696

Qy         549 GCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 608
               |    | || | |  |||||  | || || | |   || || |  |||| | |   ||||
Db     185697 GACTAGTTGGTTGGTGGGCTGGTTGGTTGGTTGGCTGGTTGGTTAGCTGGTTGGTTGGCT 185756

Qy         609 CCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 668
                 | || || | |   || || | ||||| | |   |||||  | || || | |    |
Db     185757 GGTTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGTTGGCTAGTG 185816

Qy         669 GCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTG 719
               |   || || | |   || || ||| |   ||| | |  || ||| |  ||
Db     185817 GGCTGGTTGGTTGGTTGGTTGGCTGGTTAGTTGATGGGCTGGTTGGTTTTG 185867
```

```
RESULT 69
AC116801/c
LOCUS       AC116801              160805 bp    DNA     linear   HTG 08-FEB-2003
DEFINITION  Mus musculus clone RP24-337F7, WORKING DRAFT SEQUENCE, 25 unordered
            pieces.
ACCESSION   AC116801
VERSION     AC116801.3  GI:28273484
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 160805)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-337F7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 160805)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 160805)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 8, 2003 this sequence version replaced gi:20976695.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
```

```
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
------------- Project Information
    Center project name: L25523
    Center clone name: 337_F_7
------------- Summary Statistics
    Sequencing vector: Plasmid; n/a; 100% of reads
    Chemistry: Dye-terminator Big Dye; 100% of reads
    Assembly program: Phrap; version 0.960731
    Consensus quality: 155195 bases at least Q40
    Consensus quality: 157208 bases at least Q30
    Consensus quality: 157936 bases at least Q20
    Insert size: 169000; agarose-fp
    Insert size: 158405; sum-of-contigs
    Quality coverage: 5.6 in Q20 bases; agarose-fp
    Quality coverage: 6.0 in Q20 bases; sum-of-contigs
--------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*       1     5892: contig of 5892 bp in length
*    5893     5992: gap of 100 bp
*    5993     6257: contig of 265 bp in length
*    6258     6357: gap of 100 bp
*    6358     7340: contig of 983 bp in length
*    7341     7440: gap of 100 bp
*    7441     8641: contig of 1201 bp in length
*    8642     8741: gap of 100 bp
*    8742    10219: contig of 1478 bp in length
*   10220    10319: gap of 100 bp
*   10320    12192: contig of 1873 bp in length
*   12193    12292: gap of 100 bp
*   12293    13494: contig of 1202 bp in length
*   13495    13594: gap of 100 bp
*   13595    48997: contig of 35403 bp in length
*   48998    49097: gap of 100 bp
*   49098    50944: contig of 1847 bp in length
*   50945    51044: gap of 100 bp
*   51045    52784: contig of 1740 bp in length
*   52785    52884: gap of 100 bp
*   52885    54394: contig of 1510 bp in length
*   54395    54494: gap of 100 bp
*   54495    55634: contig of 1140 bp in length
*   55635    55734: gap of 100 bp
*   55735    58329: contig of 2595 bp in length
*   58330    58429: gap of 100 bp
*   58430    60701: contig of 2272 bp in length
*   60702    60801: gap of 100 bp
*   60802    63006: contig of 2205 bp in length
*   63007    63106: gap of 100 bp
*   63107    65208: contig of 2102 bp in length
*   65209    65308: gap of 100 bp
*   65309    68610: contig of 3302 bp in length
*   68611    68710: gap of 100 bp
*   68711    71292: contig of 2582 bp in length
*   71293    71392: gap of 100 bp
*   71393    74823: contig of 3431 bp in length
*   74824    74923: gap of 100 bp
*   74924    80574: contig of 5651 bp in length
*   80575    80674: gap of 100 bp
*   80675    94613: contig of 13939 bp in length
*   94614    94713: gap of 100 bp
*   94714   109329: contig of 14616 bp in length
*  109330   109429: gap of 100 bp
*  109430   129578: contig of 20149 bp in length
*  129579   129678: gap of 100 bp
*  129679   160382: contig of 30704 bp in length
*  160383   160482: gap of 100 bp
*  160483   160805: contig of 323 bp in length.
FEATURES             Location/Qualifiers
     source          1. .160805
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-337F7"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .5892
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    5993. .6257
                     /note="assembly_fragment"
     misc_feature    6358. .7340
                     /note="assembly_fragment"
     misc_feature    7441. .8641
                     /note="assembly_fragment"
     misc_feature    8742. .10219
                     /note="assembly_fragment"
     misc_feature    10320. .12192
                     /note="assembly_fragment"
     misc_feature    12293. .13494
                     /note="assembly_fragment"
     misc_feature    13595. .48997
                     /note="assembly_fragment"
     misc_feature    49098. .50944
                     /note="assembly_fragment"
     misc_feature    51045. .52784
                     /note="assembly_fragment"
     misc_feature    52885. .54394
                     /note="assembly_fragment"
     misc_feature    54495. .55634
                     /note="assembly_fragment"
     misc_feature    55735. .58329
                     /note="assembly_fragment"
     misc_feature    58430. .60701
                     /note="assembly_fragment"
     misc_feature    60802. .63006
                     /note="assembly_fragment"
     misc_feature    63107. .65208
                     /note="assembly_fragment"
     misc_feature    65309. .68610
                     /note="assembly_fragment"
     misc_feature    68711. .71292
                     /note="assembly_fragment"
     misc_feature    71393. .74823
                     /note="assembly_fragment"
     misc_feature    74924. .80574
                     /note="assembly_fragment"
     misc_feature    80675. .94613
                     /note="assembly_fragment"
     misc_feature    94714. .109329
                     /note="assembly_fragment"
     misc_feature    109430. .129578
                     /note="assembly_fragment"

  Query Match             16.4%;  Score 163.6;  DB 2;  Length 160805;
  Best Local Similarity   74.8%;  Pred. No. 3.2e-28;
  Matches  205;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy        180 TGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGC 239
              |||||||| ||| || ||||||||     |  ||   |  ||| ||||||| |   |
Db     104300 TGCTTGCTTGCTTGCTTGCTTGCTAGCTAGCTAGCTAGCTAGCTTGCTTGCTTGCT 104241

Qy        240 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 299
              |||| ||| || |||||||||||| ||| || |||||||||||| ||| || ||||||||
Db     104240 TGCTAGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 104181

Qy        300 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 359
              |||| ||| ||  ||| ||| ||| ||| || |||||||||||| ||| || ||||||||
Db     104180 TGCTAGCTAGCTAGCTAGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 104121
```

```
Qy        360 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              |||| ||| || |||||||| ||| ||| || |||||||||||| ||| || ||||||||
Db     104120 TGCTTGCTTGCTTGCTTGCTAGCTAGCTAGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 104061

Qy        420 GAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGG 453
                  |   |   |  ||||||||||||  ||||
Db     104060 TGCTTGCTTGCTTGCTTGCTTGCTTTGGGG 104027
```

RESULT 70
AC119267
LOCUS AC119267 191953 bp DNA linear HTG 24-FEB-2003
DEFINITION Mus musculus clone RP24-322H19, WORKING DRAFT SEQUENCE, 14 unordered pieces.
ACCESSION AC119267
VERSION AC119267.3 GI:28475980
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 191953)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP24-322H19
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 191953)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 191953)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (24-FEB-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Feb 24, 2003 this sequence version replaced gi:28269582.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L25500
Center clone name: 322_H_19
---------- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 188248 bases at least Q40
Consensus quality: 189506 bases at least Q30
Consensus quality: 190207 bases at least Q20
Insert size: 188000; agarose-fp
Insert size: 190653; sum-of-contigs
Quality coverage: 9.5 in Q20 bases; agarose-fp
Quality coverage: 9.4 in Q20 bases; sum-of-contigs
----------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	6538: contig of 6538 bp in length
*	6539	6638: gap of 100 bp
*	6639	7323: contig of 685 bp in length
*	7324	7423: gap of 100 bp
*	7424	8063: contig of 640 bp in length
*	8064	8163: gap of 100 bp
*	8164	9276: contig of 1113 bp in length
*	9277	9376: gap of 100 bp
*	9377	10779: contig of 1403 bp in length
*	10780	10879: gap of 100 bp
*	10880	12442: contig of 1563 bp in length
*	12443	12542: gap of 100 bp
*	12543	13692: contig of 1150 bp in length
*	13693	13792: gap of 100 bp
*	13793	16322: contig of 2530 bp in length
*	16323	16422: gap of 100 bp
*	16423	19255: contig of 2833 bp in length
*	19256	19355: gap of 100 bp
*	19356	84569: contig of 65214 bp in length
*	84570	84669: gap of 100 bp
*	84670	112524: contig of 27855 bp in length
*	112525	112624: gap of 100 bp
*	112625	134113: contig of 21489 bp in length
*	134114	134213: gap of 100 bp
*	134214	178370: contig of 44157 bp in length
*	178371	178470: gap of 100 bp
*	178471	191953: contig of 13483 bp in length.

FEATURES Location/Qualifiers
source 1. .191953
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-322H19"
/clone_lib="RPCI-24 Male Mouse BAC"
misc_feature 1. .6538
/note="assembly_fragment
clone_end:SP6
vector_side:left"

```
     misc_feature    6639. .7323
                     /note="assembly_fragment"
     misc_feature    7424. .8063
                     /note="assembly_fragment"
     misc_feature    8164. .9276
                     /note="assembly_fragment"
     misc_feature    9377. .10779
                     /note="assembly_fragment"
     misc_feature    10880. .12442
                     /note="assembly_fragment"
     misc_feature    12543. .13692
                     /note="assembly_fragment"
     misc_feature    13793. .16322
                     /note="assembly_fragment"
     misc_feature    16423. .19255
                     /note="assembly_fragment"
     misc_feature    19356. .84569
                     /note="assembly_fragment"
     misc_feature    84670. .112524
                     /note="assembly_fragment"
     misc_feature    112625. .134113
                     /note="assembly_fragment"
     misc_feature    134214. .178370
                     /note="assembly_fragment"
     misc_feature    178471. .191953
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    54331 a  41691 c  40641 g  53989 t   1301 others
ORIGIN

  Query Match             16.3%;  Score 162.6;  DB 2;  Length 191953;
  Best Local Similarity   73.7%;  Pred. No. 5.4e-28;
  Matches  207;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        173 GTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTA 232
              ||  || |||||||| ||| || |||| |||     |  ||   |   ||| |||||||
Db     146623 GTGTGCTTGCTTGCTTGCTTGCTTGCTAGCTAGCTAGCTAGCTAGCTTGCTTGCTT 146682

Qy        233 GTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 292
              |   |   ||| ||| || |||||||||||| ||| || |||||||||||| ||| || |
Db     146683 GCTTGCTAGCTAGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 146742

Qy        293 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 352
              |||||||||||| ||| ||  ||| ||| ||| ||| || |||||||||||| ||| || |
Db     146743 GCTTGCTTGCTAGCTAGCTAGCTAGCTAGCTTGCTTGCTTGCTTGCTTGCTTGCTT 146802

Qy        353 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 412
              |||||||||| ||| || |||||||| ||| ||| || |||||||||||| ||| || |
Db     146803 GCTTGCTTGCTTGCTTGCTTGCTTGCTAGCTAGCTAGCTTGCTTGCTTGCTTGCTT 146862

Qy        413 GCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGG 453
              |||||||     |   |   |  ||||||||||||  ||||
Db     146863 GCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTTGGGG 146903


RESULT 71
AL833804
LOCUS       AL833804          .    243716 bp    DNA     linear   HTG 14-MAY-2003
DEFINITION  Mus musculus chromosome 2 clone RP23-20A6, *** SEQUENCING IN
            PROGRESS ***.
ACCESSION   AL833804
VERSION     AL833804.14  GI:30722484
KEYWORDS    HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 243716)
  AUTHORS   Griffiths,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 14, 2003 this sequence version replaced gi:29561863.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to confirm this sequence.  Sequence data
            from the whole genome shotgun alone has only been used where it has
            a phred quality of at least 30.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bM20A6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 243636 bases at least Q40
            Consensus quality: 243658 bases at least Q30
            Consensus quality: 243702 bases at least Q20
            Insert size: 243716; sum-of-contigs
            Insert size: 196320; 26.1% error; agarose-fp
            Quality coverage: 9.00x in Q20 bases; sum-of-contigs Quality
            coverage: 11.17x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   243716: contig of 243716 bp in length.
FEATURES             Location/Qualifiers
     source          1. .243716
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-20A6"
                     /clone_lib="RPCI-23"
     misc_feature    1. .243716
                     /note="assembly_fragment:00160
                     clone_end:SP6
                     vector_side:left
                     clone_end:T7
                     vector_side:right"
BASE COUNT    70331 a  54748 c  52890 g  65747 t
ORIGIN

  Query Match             16.1%;  Score 160.8;  DB 2;  Length 243716;
  Best Local Similarity   51.5%;  Pred. No. 1.4e-27;
  Matches  369;  Conservative    0;  Mismatches  347;  Indels    0;  Gaps    0;

Qy         45 GCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGG 104
              ||  ||| |  ||  | | |||       || ||   ||| | |    | | | ||   |
Db     100483 GCTGGCTGGCTAGCTAGCTGGTTGCTTGGCTTGTTGGCTGGTAGATTGGTTGGTTGGTTG 100542

Qy        105 GACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTG 164
              |   || ||   | | | | | |||   | |   ||   ||   || || | | |   |
Db     100543 GTTGGTTGGTTGGTTGGTTGGTTGGTTGATTGGCTGGTTGGTTGGTCGGTTGATTA 100602

Qy        165 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTG 224
              |  ||   |  |||  ||| ||| |||||     | |||    ||   |||||   | ||
Db     100603 ATTGGCTGGCTGGCTGGCTGGCTGGCTGGTTGCTTGGCTTGTTGGCTGGTAGGTAGGTTG 100662

Qy        225 GCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 284
              | | | | || ||   | | | |||     |  || ||| ||||||  ||  ||| | |
Db     100663 GTTGGTTGGTTGGTTGGTTGGTTGGTTAATTGACTGGCTGGCTGGCTGGCCGGCTGGTTG 100722

Qy        285 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 344
                ||||  | | ||| |   |  ||   |   | | | | | |||   | |   | |||
```

```
Db.    100723 CTTGGCTGGTTGGCTGGTAGGTAGGTAGGTAGGTTGGTTGGTTGGTTAGTTGATTGGCTG 100782

Qy        345 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 404

Db     100783 GTTGGTTGGTCGGTTGATTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTG 100842

Qy        405 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCT 464

Db     100843 GCTGGCTGGCTGGCTGGCTGGCTGGTTGCTTGGCTGGTTGGCTGGTAGGTAGGTTGATTG 100902

Qy        465 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 524

Db     100903 GTTGGTTGGTTGGTTGGCTGGCTGGCTGGCTAGTTAGTTAGTTAGTTAGTTAGTTAGTTA 100962

Qy        525 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 584

Db     100963 GTTAGTTGATTGGCTAGTTGGTTGGTCGGTTGGTTAATTGGCTGGCTGGCTGGCTGGTTG 101022

Qy        585 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 644

Db     101023 CTTGACTGGTTGGCTGGTAGGTAGGTAGGTAGGTTGGTTGGTTGGTTGGTTAGTTGATTG 101082

Qy        645 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTT 704

Db     101083 GCTGGTTGGTTGGTCGGTTGATTAATTGGCTGGCTGGCAGGCTGGCTGGCTGGCTGGCTG 101142

Qy        705 GCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAG 760

Db     101143 GCTGGCTGGCTGGCTGGCTGGCTGGCTGGTTGCTTGGCTGGCTGGCTGGTAGGTAG 101198


RESULT 72
AC124492
LOCUS       AC124492               176782 bp    DNA     linear   ROD 27-OCT-2002
DEFINITION  Mus musculus chromosome 12 clone RP23-454I23, complete sequence.
ACCESSION   AC124492
VERSION     AC124492.4  GI:24415450
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 176782)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176782)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 176782)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 176782)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Oct 27, 2002 this sequence version replaced gi:22758518.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information -----------------
            Center project name: M_BA0454I23
            ------------------------------------------------------

FEATURES             Location/Qualifiers
     source          1. .176782
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="12"
                     /clone="RP23-454I23"
BASE COUNT    55394 a  35351 c  33704 g  52333 t
ORIGIN

  Query Match           16.1%;  Score 160.6;  DB 10;  Length 176782;
  Best Local Similarity   53.8%;  Pred. No. 1.6e-27;
  Matches  331;  Conservative    0;  Mismatches  284;  Indels    0;  Gaps    0;

Qy         44 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 103

Db      40591 TGTCTGCTTCCCTGCCCGCCTGCCTGCCTGCCTGCCTGCCTGCCCGCCTGCCCGCTTCCC 40650

Qy        104 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 163

Db      40651 TGCCTGTCTGCTTCTCTGCCCGCCTGCCCGCCTGCCTGCCTGTCTGCTTCCCTGCCCGCC 40710

Qy        164 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 223

Db      40711 TGCCCGCCTGCCCGCCTGCCCGCTTCCCTGCCTGTCTGCTTCCCTGCCCGCCTGCCTGCC 40770

Qy        224 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 283

Db      40771 TGCCTGCCTGTCTGCTTCCCTGCCCGCCTGCCCGCCTGCCCGCTTGCCTGCTTCCCTGCC 40830

Qy        284 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 343

Db      40831 TGCCTGCCCGCCTGCCCGCTTGCCCGCTTGCCTGCTTCCCTGCCCGACTGCCCGCTTGCC 40890

Qy        344 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403

Db      40891 CGCTTTCCTGCCCGCCTGCCCGCCTGCCCGCCTGACCACCTGCCCGCCTGCCCGCCTGCC 40950

Qy        404 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCC 463

Db      40951 CGCCTGCCCGCCTGCCCGCCTGCCCACCTGCCCGCTTGCCTGCTTCCCTGCCCACCTGCC 41010

Qy        464 TGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 523

Db      41011 CGCCTGCCCGCTTGCCTGCTTCCCTGCCCGCCTGCCCGCTTGCCTGCTTCCCTACCCGCC 41070

Qy        524 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTA 583

Db      41071 TGCTCACCTGCCCACCTGCCCACCTGCCCACCTGCCCACCTGCCCACCTGCCCACCTGCC 41130

Qy        584 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 643

Db      41131 CGCCTGTCCGCCTGCCCGCTTCCCTGCTTCCCTGCCCACCTGCCTGCCTGCTTCCCTGCC 41190

Qy        644 GGGCTGCTGGGCTGC 658

Db      41191 TGCCTGCTTTCCTGC 41205


RESULT 73
AC098121/c
LOCUS       AC098121               231321 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-133G18, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC098121
VERSION     AC098121.7  GI:30522725
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 231321)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
```

```
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 231321)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 231321)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24957732.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GGLN
                Center clone name: CH230-133G18
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 224443 bases at least Q40
                Consensus quality: 225706 bases at least Q30
                Consensus quality: 226616 bases at least Q20
                Estimated insert size: 232595; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   229838: contig of 229838 bp in length
            *  229839   229938: gap of unknown length
            *  229939   231321: contig of 1383 bp in length.
FEATURES             Location/Qualifiers
     source          1. .231321
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-133G18"
     misc_feature    1. .3236
                     /note="wgs_contig"
     misc_feature    227078. .228053
                     /note="wgs_contig"
     misc_feature    228104. .229838
                     /note="wgs_contig"
BASE COUNT    58426 a  49296 c  52671 g  67149 t   3779 others
ORIGIN

  Query Match            16.0%;  Score 160.2;  DB 2;  Length 231321;
  Best Local Similarity  53.0%;  Pred. No. 2e-27;
  Matches  365;  Conservative    0;  Mismatches  323;  Indels    1;  Gaps    1;

Qy         69 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 128
              || ||  || ||  | | | |  || ||| | | |   |   |  ||  ||   | |
Db     182893 GGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGGTT 182834

Qy        129 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 188
              | |||  ||    |  | | |   | ||| | | ||   ||   || |||   ||  ||
Db     182833 GGCTGGCTGACTGGCTGGCTGGCTGGCTGGTGGTTGGTTGGTTGGTTGGCTGTCTGCCTG 182774

Qy        189 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 248
              | ||||   ||  |||   ||   || || |  ||| || ||| |  ||   | | | ||
Db     182773 GTTGGCTGTCTGTCTGGTTGGTTTGTTGGTCGACTGCCTGGCTGGCTGGTTGGTTTGTTG 182714

Qy        249 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 308
              |   ||| ||| ||| | ||||  |||  || | | | |||    ||   |  || ||||
Db     182713 GTTGGCTGGCTGGCTGGTTGGCTGGCTGACTGGTTGGTTGGTTGACTGATTGTCTAGCTG 182654

Qy        309 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 368
               |  | || || ||| ||||||  | | | |   | | |||   | | ||| ||| |||
Db     182653 ACTGGTTTTCTGGCTGGCTGGCTGGTTGGTTGATTTGTTGGTTGGTTGGCTGGCTGCCTG 182594

Qy        369 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 428
              |   ||| ||| ||| | |||  || | | | ||| |||||| ||| ||| |  |
Db     182593 GTTGGCTGGCTGGCTGGTTGGTTTGTTGGTTGGCTGGCTGG-CTGGTTGGTTTGTTAGTT 182535
```

```
Qy        429 GTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCT 488
              |   ||||| ||| ||| |     | | |    | ||| ||| | | ||| | |    || |
Db     182534 GGCTGGCTGGTTGGTTGGTTGATTGGCTGGCTGGCTGGCTGGTTGGTTGGTTGACTGGTT 182475

Qy        489 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 548
              |     |||   |     |||    |||| | ||  |||||   || || | |   | |
Db     182474 GTCTAACTGGCTGGTTTTCTGGCTAGCTGGTTGTCTGGCTGGCTGGTTGATTGGTTGATT 182415

Qy        549 GCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 608
              |   ||||| | |   || |  | |||||   |   |||||   |||||   |   || |
Db     182414 GGCTGGCTGGTTGATTGGTTGATTGGCTGGCTGGCTGGCTGGCTGGCTGATTGGTT 182355

Qy        609 CCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 668
               | || || | |   || || | || || | |   || || | || ||   |   ||||
Db     182354 GGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGGCT 182295

Qy        669 GCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTT 728
              |   ||||| | |   ||||| |||     ||| ||| ||| ||               |
Db     182294 GACTGGCTGGTGGGCTGGCTGACTGGCTCTTTGGTTGGTTGGTTTTGTTGTCTTTGCTAT 182235

Qy        729 GCCTGCTACTAGGGCTGCTGTGCAGCTGG 757
              ||||||| |   | ||||     ||| |
Db     182234 TCCTGCTATTGCTGTTGCTTATTTGCTTG 182206


RESULT 74
AC106600/c
LOCUS       AC106600               238112 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-227L10, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
ACCESSION   AC106600
VERSION     AC106600.4  GI:30579745
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 238112)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 238112)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 238112)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:23606102.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GLLO
                Center clone name: CH230-227L10
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 225209 bases at least Q40
                Consensus quality: 228391 bases at least Q30
                Consensus quality: 230534 bases at least Q20
                Estimated insert size: 239097; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   233995: contig of 233995 bp in length
            *  233996   234095: gap of unknown length
            *  234096   235636: contig of 1541 bp in length
            *  235637   235736: gap of unknown length
            *  235737   238112: contig of 2376 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1. .238112
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-227L10"
     misc_feature    1. .1415
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    5157. .8726
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    9883. .10719
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BZ105821"
     misc_feature    complement(231646. .232487)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BZ105821"
BASE COUNT    62132 a  48822 c  51700 g  69628 t   5830 others
ORIGIN

  Query Match             16.0%;  Score 160.2;  DB 2;  Length 238112;
  Best Local Similarity   52.6%;  Pred. No. 2e-27;
  Matches  371;  Conservative    0;  Mismatches  333;  Indels    1;  Gaps    1;

Qy          44 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 103
               || || |||   |       ||  || ||  |   |  | | | |  || ||| |  |
Db      133262 TGGTTGGTTGATTGGTTGGTTGTTTGGTTGGTTAGTTAGTTGTTTGGTTGGTTGGTTGGT 133203

Qy         104 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 163
               ||  |   | || ||| | | |  || ||| |  |   |    |   | ·|| ||| |
Db      133202 TGATTGGTTGGTTGGTTGGTTGGTTGATTGGTTGGTTGGTTGGTTGATTGATTGGTTGGT 133143

Qy         164 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 223
                 |  ||   |  || ||| | | |  || ||| |  |  |   |·|  || || |
Db      133142 TGGTTGGTAGGCTGGTTGGTTGGTTGGTTGGTTGTTTGGTTGGTTGGTTGTTTGGTTGGT 133083

Qy         224 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 283
               | ||| | |   || || | | | |  || ||| ||| | | |    || ||| ||| |
Db      133082 TGGTTGTTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTTGTTGGTTGTTTGTT 133023

Qy         284 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 343
               | | |  || ||| ||| | | | |  || ||| ||| | | | |  || ||| ||| |
Db      133022 TGTTTGGTTGGTTGTTTGGTTGTTTGGTTGGTTGATTGGTTGTTTGTTTGTTTGTTTGGT 132963

Qy         344 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403
               | | |  || ||| ||| |   | |  || ||| ||| | | | |  || ||| ||| |
Db      132962 TGGTTGGTTGGTTGGTTGGTTAGTTGTTTGGTTGTTTGGTTGGTTGGTTGGTTGGTTGAT 132903

Qy         404 CGCTGGCCTGCTTG-CTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGC 462
               | | |  || |||  ||   ||   ||  |  || ||| ||| |  |  |   |   |
Db      132902 TGGTTGTTTGTTTGTTTGGTTGGTTTGTTTGTTTGTTTGTTTGGTTGGTTGGTTGG 132843

Qy         463 CTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGT 522
               ||  || | | ||| | |   || || | || || | |    | || | || || | |
Db      132842 TTGGTTGGTTGTTTGGTTGGTTGGTTGATTGGTTGGTTGGTTAGTTGTTTGGTTGGTTGG 132783

Qy         523 AGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGT 582
               || || | || || | |   || || | || || | |   || |  | || || | |
Db      132782 TTGGTTGATTGGTTGGTTGGTTGGTTGGTTGATTGGTTGGTTGGTAGGTTGGTTGG 132723

Qy         583 AGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGT 642
               || || | || || | ||   | |  | || || | |   || || | |  || | |
Db      132722 TTGGTTGGTTGGTTGGTTGTTTAGTTGGTTGGTTGTTTGGTTGGTTGGTTGTTTGGTTGG 132663

Qy         643 AGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGC 702
               || || | || || | |   |  |  | || || | ||  || ||  || | | |||
Db      132662 TTGGTTGGTTGGTTGGTTGGTTGTTTAGTTGGTTGGTTGTTTGGTTGGTTGGTTGTTTGG 132603

Qy         703 TTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCT 747
               ||| ||| ||| | ||  ||  | |||· || |     || || |
Db      132602 TTGGTTGATTGGTTGTTTGGTTGGTTGGTTGGTTGATTGGTTGTT 132558


RESULT 75
AE014176_0/c
WPCOMMENT
  Sequence split into 4 fragments  LOCUS AE014176 Accession AE014176
     Fragment Name      Begin     End
     AE014176_0             1    110000
     AE014176_1        100001    210000
     AE014176_2        200001    310000
     AE014176_3        300001    395229
LOCUS       AE014176               395229 bp    DNA     linear   ROD 06-AUG-2002
DEFINITION  Mus musculus piebald deletion region section 4 of 11 of the
            complete sequence.
ACCESSION   AE014176 AE013600
VERSION     AE014176.1  GI:22128031
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 395229)
  AUTHORS   Peterson,K., King,B., Hagge-Greenberg,A., Roix,J., Bult,C. and
            O'Brien,T.
  TITLE     Functional and comparative genomic analysis of the piebald deletion
            region of mouse chromosome 14
  JOURNAL   Genomics 80 (2), 172 (2002)
  MEDLINE   22150871
   PUBMED   12160731
REFERENCE   2  (bases 1 to 395229)
  AUTHORS   Adams,M. and Mural,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAY-2002) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     The strings of n's in this record represent gaps between contigs,
            and the length of each string corresponds to the length of the gap
            except where there are 20 n's.  A string of 20 n's represents a gap
            of unknown length.
FEATURES             Location/Qualifiers
     source          1. .395229
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="14"
                     /map="Mit164-Mit165"
     gene            <109204. .>207970
                     /gene="Scel"
     mRNA            <109204. .>207970
                     /gene="Scel"
                     /product="sciellin"
                     /note="partial transcript sequence derived from alignment
                     to Mus musculus NCBI RefSeq sequence NM_022886"
     mRNA            <245271. .>299392
                     /product="TC255155"
                     /note="partial transcript derived from alignment to
                     tentative consensus sequence TC200980, from TIGR Mouse
                     Gene Index (v. 5.0); developmental expression detected by
                     RT-PCR using panel of embryonic RNAs  (E8.5, E10.5, E12.5,
                     E14.5, E16.5 and E18.5) from C57BL/6J mice"
BASE COUNT   101118 a  74785 c  76350 g 109777 t  33199 others
ORIGIN

  Query Match             16.0%;  Score 160;  DB 10;  Length 110000;
  Best Local Similarity   54.9%;  Pred. No. 2.3e-27;
  Matches  336;  Conservative    0;  Mismatches  275;  Indels    1;  Gaps    1;

Qy          82 CTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGA 141
               || |||| |  |||  ||||  |    |  | |   ||  ||  |||||   ||  ||||
Db      60852 CTGGCTGACTGGCTGACTGACTGACTGGCTGACTGACTGACTGACTGGCTGACTGACTGA 60793
```

```
Qy        142 GGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTG 201
              ||   | |||   ||  |||   |    || |||   || ||| |||| |   || |
Db      60792 CTGGCTGGCTGGCTGACTGACTGGCTGACTGGTTGGCTGACTGGCTGGCTGACTGACTGG 60733

Qy        202 CTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTG 261
              ||||  |    | |    |||||||  ||    ||    ||  |||||   ||  ||
Db      60732 CTGACTGACTGGCTGACTGGCTGGCTGACTGACTGGCTGACTGACTGGCTGACTGACTGA 60673

Qy        262 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 321
              ||  ||| |   ||  ||  || ||||||  ||| ||| ||| ||||||  |||  || |
Db      60672 CTGACTGACTGACTGACTGACTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGACTGG 60613

Qy        322 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 381
              || ||||||  ||| ||| ||| ||||||  ||| ||| |||  ||| |   ||  ||
Db      60612 CTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGACTGACTGACTGACTGA 60553

Qy        382 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTG 441
              ||  |||||  ||| ||| ||| ||||||  ||| |||||  ||   |  ||   ||| |
Db      60552 CTGACTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGACTGGCTGGCTGGCTGGCTGG 60493

Qy        442 CTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAG 501
              || ||| |   ||| |   | ||| ||||  |  |||  |   | |||   | |||   |
Db      60492 CTGGCTGA-CTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTG 60434

Qy        502 TAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAG 561
              |||||  | |||   |  | ||| | |||||   |   |||||. | |||   |
Db      60433 ACTGGCTGGCTGACTGACTGGCTGACTGGTTGGCTGGCTGGCTGGCTGGCTGACTGGCTG 60374

Qy        562 TAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAG 621
              | ||   | |||   |   | |||   | |||   |   | ||   | |||  | |
Db      60373 ACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGGTTG 60314

Qy        622 TAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAG 681
              |||||  |||||  |  |||||  | |||  |  |||||  | |||
Db      60313 GCTGGCTGGCTGGCTGACTGACTGGCTGACTGACTGACTGACTGGCTGACTGACTGGCTA 60254

Qy        682 TAGGGCTGCCTG 693
              || || |||
Db      60253 ACTGGTTGGCTG 60242
```

RESULT 76
AC114585/c
LOCUS AC114585 189576 bp DNA linear HTG 17-MAR-2003
DEFINITION Mus musculus clone RP23-296J16, WORKING DRAFT SEQUENCE, 10 unordered pieces.
ACCESSION AC114585
VERSION AC114585.7 GI:28975893
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 189576)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP23-296J16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 189576)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 189576)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (17-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 17, 2003 this sequence version replaced gi:28412057.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L23692
Center clone name: 296_J_16
---------- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 186712 bases at least Q40
Consensus quality: 187848 bases at least Q30
Consensus quality: 188381 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 188676; sum-of-contigs
Quality coverage: 8.0 in Q20 bases; agarose-fp
Quality coverage: 8.2 in Q20 bases; sum-of-contigs
----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 47737: contig of 47737 bp in length
* 47738 47837: gap of 100 bp
* 47838 48894: contig of 1057 bp in length
* 48895 48994: gap of 100 bp
* 48995 49945: contig of 951 bp in length
* 49946 50045: gap of 100 bp
* 50046 51530: contig of 1485 bp in length

```
            *    51531    51630: gap of 100 bp
            *    51631    57122: contig of 5492 bp in length
            *    57123    57222: gap of 100 bp
            *    57223    66812: contig of 9590 bp in length
            *    66813    66912: gap of 100 bp
            *    66913   136223: contig of 69311 bp in length
            *   136224   136323: gap of 100 bp
            *   136324   146728: contig of 10405 bp in length
            *   146729   146828: gap of 100 bp
            *   146829   185215: contig of 38387 bp in length
            *   185216   185315: gap of 100 bp
            *   185316   189576: contig of 4261 bp in length.
FEATURES             Location/Qualifiers
     source          1. .189576
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-296J16"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .47737
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    47838. .48894
                     /note="assembly_fragment"
     misc_feature    48995. .49945
                     /note="assembly_fragment"
     misc_feature    50046. .51530
                     /note="assembly_fragment"
     misc_feature    51631. .57122
                     /note="assembly_fragment"
     misc_feature    57223. .66812
                     /note="assembly_fragment"
     misc_feature    66913. .136223
                     /note="assembly_fragment"
     misc_feature    136324. .146728
                     /note="assembly_fragment"
     misc_feature    146829. .185215
                     /note="assembly_fragment"
     misc_feature    185316. .189576
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    58607 a  38256 c  38652 g  53160 t    901 others
ORIGIN

  Query Match             16.0%;  Score 160;  DB 2;  Length 189576;
  Best Local Similarity   54.9%;  Pred. No. 2.2e-27;
  Matches  336;  Conservative    0;  Mismatches  275;  Indels    1;  Gaps    1;

Qy          82 CTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGA 141
               || |||| |  |||  |||| |    |  | |   ||  ||  |||||   ||  ||||
Db      172936 CTGGCTGACTGGCTGACTGACTGACTGGCTGACTGACTGACTGACTGGCTGACTGACTGA 172877

Qy         142 GGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTG 201
                 ||   |  |||   ||  |||    |    || |||   || ||| |||| |   || |
Db      172876 CTGGCTGGCTGGCTGACTGACTGGCTGACTGGTTGGCTGACTGGCTGGCTGACTGACTGG 172817

Qy         202 CTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTG 261
               ||||  |    |  |    |||||||  ||    ||    ||  |||||   ||  ||
Db      172816 CTGACTGACTGGCTGACTGGCTGGCTGACTGACTGGCTGACTGACTGGCTGACTGACTGA 172757

Qy         262 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 321
               ||  ||| |   ||  ||  || ||||||  ||| ||| ||| ||||||  |||  || |
Db      172756 CTGACTGACTGACTGACTGACTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGACTGACTGG 172697

Qy         322 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 381
               || ||||||  ||| ||| ||| ||||||  ||| ||| |||  ||| |   ||  ||
Db  .   172696 CTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGACTGACTGACTGACTGA 172637

Qy         382 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTG 441
               ||  |||||  ||| ||| ||| ||||||  ||| |||||  ||    |  ||   ||| |
Db      172636 CTGACTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGACTGGCTGGCTGGCTGGCTGG 172577

Qy         442 CTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAG 501
               || ||| |   ||| |   | ||| ||||  |  |||  |   | |||   | |||   |
Db      172576 CTGGCTGA-CTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTG 172518

Qy         502 TAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAG 561
                 | |||   |||||   |   |||||   || ||   |   |||||   |||||  | |
Db      172517 ACTGACTGGCTGGCTGACTGACTGGCTGACTGGTTGGCTGGCTGGCTGGCTGGCTGACTG 172458

Qy         562 TAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAG 621
                 | ||    | |||   |   | |||   | |||   |   | ||    | ||| | |
Db      172457 GCTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGGTTG 172398

Qy         622 TAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAG 681
                 |||||   |||||   |   |||||   | |||   |   |||||   | |||
Db      172397 GCTGGCTGGCTGGCTGACTGACTGGCTGACTGACTGACTGACTGGCTGACTGACTGGCTA 172338

Qy         682 TAGGGCTGCCTG 693
                 || || |||
Db      172337 ACTGGTTGGCTG 172326


RESULT 77
AC108277_1/c
WPCOMMENT
  Sequence split into 5 fragments  LOCUS AC108277 Accession AC108277
     Fragment Name      Begin     End
     AC108277_0             1    110000
     AC108277_1        100001    210000
     AC108277_2        200001    310000
     AC108277_3        300001    410000
     AC108277_4        400001    492135
Continuation (2 of 5) of AC108277 from base 100001 (AC108277 Rattus norvegicus clone

  Query Match             16.0%;  Score 159.8;  DB 2;  Length 110000;
  Best Local Similarity   67.0%;  Pred. No. 2.6e-27;
  Matches  227;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy          91 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 150
               |||   | |   || |  |   ||||||||||  ||  ||||||||||      | | |
Db      69644 CTGTAGGTTAGTGGTAGTGCCTGCCTGCTTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGC 69585

Qy         151 AGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 210
                 ||||||||||      | | |   ||||||||||  ||| ||||||||||      |
Db      69584 TTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGC 69525

Qy         211 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 270
               | |   | ||||  || |||  |   | |||||  ||  ||  |||  |||  |||  ||| ||
Db      69524 CTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCCTGCCTGCTTGC 69465

Qy         271 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 330
               |||| |||||||  ||| ||||||  |||||||  |||  |||| | |||  |||  ||  ||
Db      69464 CTGCCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTTCCTGCCTGCCTGCCTGC 69405

Qy         331 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 390
               |||| ||| |||  ||  || ||| |||||||  ||| ||||||||| |||| ||  ||
Db      69404 CTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGC 69345

Qy         391 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 429
               |||| |||||||  ||| ||||||||||      ||||
Db      69344 CTGCCTGCTTGCCTGCTTGCCTGCTTGCTTGTCAAACAG 69306


RESULT 78
AC132758
LOCUS       AC132758              238255 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-101A12, *** SEQUENCING IN PROGRESS
            ***, 4 unordered pieces.
ACCESSION   AC132758
VERSION     AC132758.3  GI:25073372
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 238255)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 238255)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 238255)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:23195993.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KBYE
                Center clone name: CH230-101A12
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 225270 bases at least Q40
                Consensus quality: 227698 bases at least Q30
                Consensus quality: 229450 bases at least Q20
                Estimated insert size: 226046; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    15508: contig of 15508 bp in length
            *    15509    15608: gap of unknown length
            *    15609   224342: contig of 208734 bp in length
            *   224343   224442: gap of unknown length
            *   224443   225652: contig of 1210 bp in length
            *   225653   225752: gap of unknown length
            *   225753   238255: contig of 12503 bp in length.
FEATURES             Location/Qualifiers
     source          1. .238255
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-101A12"
     misc_feature    1. .1480
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    1361. .2115
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BH306826"
     misc_feature    15609. .16703
                     /note="wgs_contig"
     misc_feature    38908. .40937
                     /note="wgs_contig"
     misc_feature    41382. .43075
                     /note="wgs_contig"
     misc_feature    222951. .224342
                     /note="wgs_contig"
BASE COUNT    63779 a  53533 c  52534 g  60451 t   7958 others
ORIGIN

  Query Match            16.0%;  Score 159.8;  DB 2;  Length 238255;
  Best Local Similarity  67.0%;  Pred. No. 2.4e-27;
  Matches  227;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy         91 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 150
              |||   | |  || |  | ||||||||||  ||  ||||||||||     | | |
Db      26672 CTGTAGGTTAGTGGTAGTGCCTGCCTGCTTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGC 26731
```

```
Qy        151 AGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 210
              ||||||||||        | | |   ||||||||||  ||| ||||||||||      |
Db      26732 TTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGC 26791

Qy        211 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 270
              | |   | ||||  || |||  |   | |||||  ||  || ||| ||| |||  ||| ||
Db      26792 CTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCCTGCCTGCTTGC 26851

Qy        271 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 330
              |||| |||||||  ||| ||||||  |||||||  ||| |||| | ||| |||  ||  ||
Db      26852 CTGCCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTTCCTGCCTGCCTGCCTGC 26911

Qy        331 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 390
              |||| ||| |||  ||  || ||| |||||||  ||| ||||||||| |||| ||  ||
Db      26912 CTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGC 26971

Qy        391 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 429
              |||| |||||||  ||| ||||||||||      ||||
Db      26972 CTGCCTGCTTGCCTGCTTGCCTGCTTGCTTGTCAAACAG 27010
```

RESULT 79
AC116206
LOCUS AC116206 240200 bp DNA linear HTG 09-NOV-2002
DEFINITION Rattus norvegicus clone CH230-10104, WORKING DRAFT SEQUENCE, 5 unordered pieces.
ACCESSION AC116206
VERSION AC116206.4 GI:24818933
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 240200)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 240200)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 240200)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (09-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 9, 2002 this sequence version replaced gi:23101726.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GMPU
Center clone name: CH230-10104
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 227436 bases at least Q40
Consensus quality: 229834 bases at least Q30
Consensus quality: 231317 bases at least Q20
Estimated insert size: 232522; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 62473: contig of 62473 bp in length
* 62474 62573: gap of unknown length
* 62574 236709: contig of 174136 bp in length
* 236710 236809: gap of unknown length
* 236810 237884: contig of 1075 bp in length
* 237885 237984: gap of unknown length
* 237985 239022: contig of 1038 bp in length
* 239023 239122: gap of unknown length
* 239123 240200: contig of 1078 bp in length.

```
FEATURES             Location/Qualifiers
     source          1. .240200
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-10104"
     misc_feature    1. .1428
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    2029. .2702
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH288965"
     misc_feature    complement(53971. .54130)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH288967"
BASE COUNT    64992 a  53941 c  53130 g  60171 t   7966 others
ORIGIN

  Query Match             16.0%;  Score 159.8;  DB 2;  Length 240200;
  Best Local Similarity   67.0%;  Pred. No. 2.4e-27;
  Matches  227;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy         91 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 150
              |||   | |   || |  |   |||||||||| ||  ||||||||||      | | |
Db      11706 CTGTAGGTTAGTGGTAGTGCCTGCCTGCTTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGC 11765

Qy        151 AGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 210
              ||||||||||      | | |   |||||||||| ||| ||||||||||      |
Db      11766 TTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGC 11825

Qy        211 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 270
              | |   | ||||  || |||  |   | ||||| ||  || ||| ||| |||  ||| ||
Db      11826 CTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCCTGCCTGCTTGC 11885

Qy        271 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 330
              |||| |||||||  ||| ||||||  |||||||  ||| |||| | ||| |||  ||  ||
Db      11886 CTGCCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTTCCTGCCTGCCTGCCTGC 11945

Qy        331 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 390
              |||| ||| |||  ||  || ||| |||||||  ||| ||||||||| |||| ||  ||
Db      11946 CTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGC 12005

Qy        391 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 429
              |||| |||||||  ||| ||||||||||      ||||
Db      12006 CTGCCTGCTTGCCTGCTTGCCTGCTTGCTTGTCAAACAG 12044


RESULT 80
AC103570
LOCUS       AC103570              242974 bp    DNA     linear   HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-10M24, WORKING DRAFT SEQUENCE.
ACCESSION   AC103570
VERSION     AC103570.9  GI:24818028
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 242974)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., .
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 242974)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 242974)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 9, 2002 this sequence version replaced gi:22772757.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
```

```
            Center project name: GDRO
            Center clone name: CH230-10M24
          ---------- Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 231206 bases at least Q40
            Consensus quality: 232201 bases at least Q30
            Consensus quality: 232817 bases at least Q20
            Estimated insert size: 239753; sum-of-contigs estimation
            Quality coverage: 9x in Q20 bases; sum-of-contigs estimation
          ----------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 1 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *       1   242974: contig of 242974 bp in length.
FEATURES             Location/Qualifiers
     source          1. .242974
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-10M24"
     misc_feature    1. .2423
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    3420. .3992
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH308241"
     misc_feature    238677. .242974
                     /note="wgs_contig"
BASE COUNT    65335 a  52835 c  51454 g  63642 t   9708 others
ORIGIN

  Query Match             16.0%;  Score 159.8;  DB 2;  Length 242974;
  Best Local Similarity   67.0%;  Pred. No. 2.4e-27;
  Matches  227;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy          91 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 150
               |||   | |   || |  |   ||||||||||  ||  ||||||||||     | | |
Db      194627 CTGTAGGTTAGTGGTAGTGCCTGCCTGCTTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGC 194686

Qy         151 AGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 210
                ||||||||||      | | |   ||||||||||  ||| ||||||||||      |
Db      194687 TTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGC 194746

Qy         211 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 270
               | |   | ||||  || |||  |   | |||||  ||  || ||| ||| |||  ||| ||
Db      194747 CTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCCTGCCTGCTTGC 194806

Qy         271 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 330
               |||| |||||||  ||| ||||||  |||||||  ||| |||| | ||| |||  ||  ||
Db      194807 CTGCCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTTCCTGCCTGCCTGCCTGC 194866

Qy         331 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 390
               |||| ||| |||  ||  || ||| |||||||  ||| ||||||||||| |||| ||  ||
Db      194867 CTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGC 194926

Qy         391 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 429
               |||| |||||||  ||| ||||||||||      ||||
Db      194927 CTGCCTGCTTGCCTGCTTGCCTGCTTGCTTGTCAAACAG 194965


RESULT 81
AC116072/c
LOCUS       AC116072              244301 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-249G11, *** SEQUENCING IN PROGRESS
            ***, 4 unordered pieces.
ACCESSION   AC116072
VERSION     AC116072.5  GI:25078938
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 244301)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 244301)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 244301)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
```

```
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:22857224.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GNEY
                Center clone name: CH230-249G11
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 205700 bases at least Q40
                Consensus quality: 209464 bases at least Q30
                Consensus quality: 211942 bases at least Q20
                Estimated insert size: 222435; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   239389: contig of 239389 bp in length
            *  239390   239489: gap of unknown length
            *  239490   240661: contig of 1172 bp in length
            *  240662   240761: gap of unknown length
            *  240762   242103: contig of 1342 bp in length
            *  242104   242203: gap of unknown length
            *  242204   244301: contig of 2098 bp in length.
FEATURES             Location/Qualifiers
     source          1. .244301
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-249G11"
     misc_feature    1. .1076
                     /note="wgs_contig"
     misc_feature    3053. .4222
                     /note="wgs_contig"
     misc_feature    83375. .84860
                     /note="wgs_contig"
     misc_feature    238099. .239389
                     /note="wgs_contig"
BASE COUNT    59847 a  48345 c  47032 g  58326 t  30751 others
ORIGIN

  Query Match             16.0%;  Score 159.8;  DB 2;  Length 244301;
  Best Local Similarity   67.0%;  Pred. No. 2.4e-27;
  Matches  227;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy          91 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 150
               |||   | |   || |  |   ||||||||||  ||  ||||||||||      | | |
Db      132258 CTGTAGGTTAGTGGTAGTGCCTGCCTGCTTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGC 132199

Qy          151 AGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 210
                ||||||||||      | | |   ||||||||||  ||| ||||||||||      |
Db      132198 TTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGC 132139

Qy          211 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 270
                | |   | ||||  || |||  |   | |||||  ||  ||  ||| |||  |||  || ||
Db      132138 CTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCCTGCCTGCTTGC 132079

Qy          271 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 330
                ||||  |||||||  ||| ||||||  |||||||  ||| ||||  | |||  |||  ||  ||
Db      132078 CTGCCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTTCCTGCCTGCCTGCCTGC 132019

Qy          331 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 390
                ||||  |||  |||  ||  ||  |||  |||||||  |||  ||||||||||  ||||  ||  ||
Db      132018 CTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGC 131959

Qy          391 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 429
                ||||  |||||||  |||  ||||||||||      ||||
Db      131958 CTGCCTGCTTGCCTGCTTGCCTGCTTGCTTGTCAAACAG 131920


RESULT 82
AC106240/c
LOCUS       AC106240              251656 bp    DNA     linear   HTG 22-SEP-2002
DEFINITION  Rattus norvegicus clone CH230-129E1, *** SEQUENCING IN PROGRESS
            ***, 6 unordered pieces.
ACCESSION   AC106240
VERSION     AC106240.3  GI:23268191
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 251656)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
```

```
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 251656)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 251656)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 22, 2002 this sequence version replaced gi:21731477.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequening reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/).  As a result, the
            sequence may extend beyond the ends of the clone and there may be
            contigs that consist entirely of whole genome shotgun sequence
            reads.  Both end sequences and whole genome shotgun sequence only
            contigs will be indicated in the feature table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GKEV
                Center clone name: CH230-129E1
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 226971 bases at least Q40
                Consensus quality: 228331 bases at least Q30
                Consensus quality: 229024 bases at least Q20
                Estimated insert size: 260341; sum-of-contigs estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    71565: contig of 71565 bp in length
            *    71566    71665: gap of unknown length
            *    71666   165635: contig of 93970 bp in length
            *   165636   165735: gap of unknown length
            *   165736   246680: contig of 80945 bp in length
            *   246681   246780: gap of unknown length
            *   246781   247856: contig of 1076 bp in length
            *   247857   247956: gap of unknown length
            *   247957   249454: contig of 1498 bp in length
            *   249455   249554: gap of unknown length
            *   249555   251656: contig of 2102 bp in length.
FEATURES             Location/Qualifiers
     source          1. .251656
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-129E1"
     misc_feature    1. .1121
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    21157. .21283
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH286379"
     misc_feature    71666. .73022
                     /note="wgs_contig"
BASE COUNT    62675 a  52099 c  51098 g  63549 t  22235 others
ORIGIN

  Query Match             15.9%;  Score 159.2;  DB 2;  Length 251656;
  Best Local Similarity   54.7%;  Pred. No. 3.3e-27;
  Matches  359;  Conservative    0;  Mismatches  293;  Indels    4;  Gaps    2;

Qy         81 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 140
              | | | ||    | | | ||   ||   || ||   | | | |   |||   | ||| |
Db     223706 GTTGGTTGTTTGGTTGGTTGTATGGTTGGTTGGTGGGTTGGTTGATTGGTGGGGTTGGTT 223647

Qy        141 AGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTT 200
               |   |   |  || ||| || |   |   |  || ||| | | | |  || ||
Db     223646 GGTTGGTGGATTGGTTGGTTGGTGTGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 223587

Qy        201 GCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTT 260
              | ||    |   |   || || ||| ||| | |  ||| || | | | |   | ||| ||
Db     223586 GGTGGATTGGTGGGTTGGTTGGTGGGTTGGTTGGTGGGTTGGTTGGTTGGTGGGTTGGTT 223527

Qy        261 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 320
              | |   |||  || ||| ||| | | |||  || ||| | | | | | |  || ||| |
Db     223526 GGTTTGTGGGTTGGTTGGTTGGTTGGTGGGTTGGTTGGTGGGTTGGTTGGTTGGTTGGTG 223467

Qy        321 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 380
              | | | |||  || |||  || | | |||  || ||| ||| | | | |  || ||| |
Db     223466 GGTTGGTGGGTTGGTTGGATGGTTGGTGGGTTGGTTGGTTGGTGGGTTGGTTGGTTGGTG 223407

Qy        381 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTT 440
              | | | |||  || ||| | | | | |||  || ||| || |   |   || || ||
Db     223406 GGTTGGTGGGTTGGTTGGTTGGGTTGGTGGGTTGGTTGGTTGGTTGGTTGGTTGGTT 223347

Qy        441 GCTTGCTAAG-GGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCT 499
              | ||| |  |  ||  ||| ||   |   | ||| ||| | |  ||| || |||| || |
Db     223346 GGTTGGTGGGTTGGTTGGTTCGTTGGTGGGTTGGTTTGGTTGGTGGGTTGGTGGGTTGGT 223287

Qy        500 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 559
              |  ||| || | || || | |   || || |||| || | |   ||  | | || || | |
Db     223286 TGGTGGGTTGGTTGGTTGGTTGGTTGGTGGGTTGGTGGGTTGGTGGGTTGGTTGGT 223227

Qy        560 AGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCT 619
              |  ||| |  | ||  |   || ||  || | || || | |   || |  | || || |
Db     223226 TGGTGGGTTGGTTGGTGG---GTTGGTTTGGTTGGTTGGTTGGTTGGTTGGTTGGT 223170

Qy        620 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 679
              |  ||| || | || || | |   || || | || || | |  ||| || | || || |
Db     223169 TGGTGGGTTGGTTGGTTGGTTGGTTGGTTGGTTGATTGGTGGGTTGGTTGGTTGGT 223110

Qy        680 AGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCT 735
              |   || ||   | | | | | ||| ||| ||| | |  |||  | |||  || |
Db     223109 GGATTGGTTGGTGGGTTGGTGGGTTGGTTGATTGGTTGGTGGGTTGGTTGGTTGAT 223054


RESULT 83
AC122561/c
LOCUS       AC122561              208994 bp    DNA     linear   HTG 17-MAR-2003
DEFINITION  Mus musculus clone RP24-97G2, WORKING DRAFT SEQUENCE, 10 unordered
            pieces.
ACCESSION   AC122561
VERSION     AC122561.3  GI:28975915
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
```

```
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 208994)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-97G2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 208994)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 208994)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 17, 2003 this sequence version replaced gi:28882445.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L26523
                Center clone name: 97_G_2
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 205816 bases at least Q40
                Consensus quality: 206724 bases at least Q30
                Consensus quality: 207402 bases at least Q20
                Insert size: 183000; agarose-fp
                Insert size: 208094; sum-of-contigs
                Quality coverage: 9.8 in Q20 bases; agarose-fp
                Quality coverage: 8.7 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2678: contig of 2678 bp in length
            *     2679     2778: gap of 100 bp
            *     2779     7560: contig of 4782 bp in length
            *     7561     7660: gap of 100 bp
            *     7661    17565: contig of 9905 bp in length
            *    17566    17665: gap of 100 bp
            *    17666    31160: contig of 13495 bp in length
            *    31161    31260: gap of 100 bp
            *    31261    53826: contig of 22566 bp in length
            *    53827    53926: gap of 100 bp
            *    53927    84289: contig of 30363 bp in length
            *    84290    84389: gap of 100 bp
            *    84390   109590: contig of 25201 bp in length
            *   109591   109690: gap of 100 bp
            *   109691   140213: contig of 30523 bp in length
            *   140214   140313: gap of 100 bp
            *   140314   172070: contig of 31757 bp in length
            *   172071   172170: gap of 100 bp
            *   172171   208994: contig of 36824 bp in length.
FEATURES             Location/Qualifiers
     source          1. .208994
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-97G2"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .2678
                     /note="assembly_fragment"
     misc_feature    2779. .7560
                     /note="assembly_fragment"
     misc_feature    7661. .17565
                     /note="assembly_fragment"
     misc_feature    17666. .31160
                     /note="assembly_fragment"
     misc_feature    31261. .53826
                     /note="assembly_fragment"
     misc_feature    53927. .84289
                     /note="assembly_fragment"
     misc_feature    84390. .109590
                     /note="assembly_fragment"
     misc_feature    109691. .140213
                     /note="assembly_fragment"
     misc_feature    140314. .172070
                     /note="assembly_fragment"
     misc_feature    172171. .208994
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    61737 a  42018 c  43038 g  61299 t    902 others
ORIGIN

  Query Match             15.9%;  Score 158.8;  DB 2;  Length 208994;
  Best Local Similarity   68.3%;  Pred. No. 4.2e-27;
  Matches  220;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy        100 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCT 159
```

```
                ||  | | || ||  |   |||    || |  ||||||||     |   ||  || ||||
Db      46080 GAATGAAAAGCAGATCCATTTGTCTACTTGTTTGCTTGCTTGCTTGCTTGTTTGCTTGCT 46021

Qy        160 TGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGC 219
              | ||     |   ||  || |||||||| ||| || ||||||||     |       |
Db      46020 TACTTGTTTGCTTGTTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTCACTTGCT 45961

Qy        220 TGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 279
              |||| |||   | |   |  |||| ||| || || ||||||||| ||| || ||||||||
Db      45960 TGCTCGCTCCTTTGCTCGCTTGCTTGCTTGCTTGTTTGCTTGCTTGCTTGCTTGCTTGCT 45901

Qy        280 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 339
              |||| ||| || ||||||||||| ||| || || ||||||||| ||| || || |||||
Db      45900 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGTTTGCTTGCTTGCTTGCTTGTTTGCT 45841

Qy        340 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 399
              |||| | | || || ||||||| | ||| || |||||||||||| ||| || ||||||||
Db      45840 TGCTTGTTTGCTTGTTTGCTTGTTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 45781

Qy        400 TGCTCGCTGGCCTGCTTGCTGA 421
              |||| ||| || ||||| || |
Db      45780 TGCTTGCTTGCTTGCTTCCTCA 45759


RESULT 84
AL772377
LOCUS       AL772377               217837 bp    DNA     linear   ROD 08-OCT-2002
DEFINITION  Mouse DNA sequence from clone RP23-141B15 on chromosome 2, complete
            sequence.
ACCESSION   AL772377
VERSION     AL772377.6  GI:23893026
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 217837)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 12, 2002 this sequence version replaced gi:23395436.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-141B15 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .217837
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-141B15"
                     /clone_lib="RPCI-23"
BASE COUNT    65927 a  46456 c  45477 g  59977 t
ORIGIN

  Query Match           15.9%;  Score 158.8;  DB 10;  Length 217837;
  Best Local Similarity   68.3%;  Pred. No. 4.2e-27;
  Matches  220;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy        100 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCT 159
              ||  | | || ||  |   |||    || |  ||||||||     |   ||  || ||||
Db      37190 GAATGAAAAGCAGATCCATTTGTCTACTTGTTTGCTTGCTTGCTTGCTTGTTTGCTTGCT 37249

Qy        160 TGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGC 219
              | ||     |   ||  || |||||||| ||| || ||||||||     |       |
Db      37250 TACTTGTTTGCTTGTTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTCACTTGCT 37309

Qy        220 TGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 279
              |||| |||   | |   |  |||| ||| || || ||||||||| ||| || ||||||||
Db      37310 TGCTCGCTCCTTTGCTCGCTTGCTTGCTTGCTTGTTTGCTTGCTTGCTTGCTTGCTTGCT 37369

Qy        280 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 339
              |||| ||| || ||||||||||| ||| || || ||||||||| ||| || || |||||
Db      37370 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGTTTGCTTGCTTGCTTGCTTGTTTGCT 37429

Qy        340 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 399
              |||| | | || || |||||||| | ||| || |||||||||||| ||| || ||||||||
Db      37430 TGCTTGTTTGCTTGTTTGCTTGTTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 37489

Qy        400 TGCTCGCTGGCCTGCTTGCTGA 421
              |||| ||| || ||||| || |
Db      37490 TGCTTGCTTGCTTGCTTCCTCA 37511


RESULT 85
AC044848
LOCUS       AC044848               179192 bp    DNA     linear   HTG 09-MAY-2001
DEFINITION  Mus musculus chromosome 11 clone RP23-129L22 map 11, WORKING DRAFT
            SEQUENCE, 70 unordered pieces.
ACCESSION   AC044848
VERSION     AC044848.3  GI:12658164
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 179192)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 11, clone RP23-129L22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179192)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
```

```
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 3, 2001 this sequence version replaced gi:10717286.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9178
                Center clone name: 129_L_22
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 7% of reads
                Sequencing vector: Plasmid; n/a; 93% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 151306 bases at least Q40
                Consensus quality: 164443 bases at least Q30
                Consensus quality: 169708 bases at least Q20
                Insert size: 230000; agarose-fp
                Insert size: 172292; sum-of-contigs
                Quality coverage: 2.4 in Q20 bases; agarose-fp
                Quality coverage: 3.2 in Q20 ba.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 70 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      399: contig of 399 bp in length
            *      400      499: gap of 100 bp
            *      500     1142: contig of 643 bp in length
            *     1143     1242: gap of 100 bp
            *     1243     1653: contig of 411 bp in length
            *     1654     1753: gap of 100 bp
            *     1754     2376: contig of 623 bp in length
            *     2377     2476: gap of 100 bp
            *     2477     3095: contig of 619 bp in length
            *     3096     3195: gap of 100 bp
            *     3196     4176: contig of 981 bp in length
            *     4177     4276: gap of 100 bp
            *     4277     5264: contig of 988 bp in length
            *     5265     5364: gap of 100 bp
            *     5365     6039: contig of 675 bp in length
            *     6040     6139: gap of 100 bp
            *     6140     6817: contig of 678 bp in length
            *     6818     6917: gap of 100 bp
            *     6918     7954: contig of 1037 bp in length
            *     7955     8054: gap of 100 bp
            *     8055     8811: contig of 757 bp in length
            *     8812     8911: gap of 100 bp
            *     8912     9775: contig of 864 bp in length
            *     9776     9875: gap of 100 bp
            *     9876    10921: contig of 1046 bp in length
            *    10922    11021: gap of 100 bp
            *    11022    12232: contig of 1211 bp in length
            *    12233    12332: gap of 100 bp
            *    12333    13351: contig of 1019 bp in length
            *    13352    13451: gap of 100 bp
            *    13452    14326: contig of 875 bp in length
            *    14327    14426: gap of 100 bp
            *    14427    15792: contig of 1366 bp in length
            *    15793    15892: gap of 100 bp
            *    15893    16562: contig of 670 bp in length
            *    16563    16662: gap of 100 bp
            *    16663    17540: contig of 878 bp in length
            *    17541    17640: gap of 100 bp
            *    17641    18844: contig of 1204 bp in length
            *    18845    18944: gap of 100 bp
            *    18945    20507: contig of 1563 bp in length
            *    20508    20607: gap of 100 bp
            *    20608    21903: contig of 1296 bp in length
            *    21904    22003: gap of 100 bp
            *    22004    23064: contig of 1061 bp in length
            *    23065    23164: gap of 100 bp
            *    23165    24725: contig of 1561 bp in length
            *    24726    24825: gap of 100 bp
            *    24826    26288: contig of 1463 bp in length
            *    26289    26388: gap of 100 bp
            *    26389    28040: contig of 1652 bp in length
            *    28041    28140: gap of 100 bp
            *    28141    28975: contig of 835 bp in length
            *    28976    29075: gap of 100 bp
            *    29076    29960: contig of 885 bp in length
            *    29961    30060: gap of 100 bp
            *    30061    31316: contig of 1256 bp in length
            *    31317    31416: gap of 100 bp
            *    31417    32758: contig of 1342 bp in length
            *    32759    32858: gap of 100 bp
            *    32859    34220: contig of 1362 bp in length
            *    34221    34320: gap of 100 bp
            *    34321    35754: contig of 1434 bp in length
            *    35755    35854: gap of 100 bp
            *    35855    37185: contig of 1331 bp in length
            *    37186    37285: gap of 100 bp
            *    37286    39443: contig of 2158 bp in length
            *    39444    39543: gap of 100 bp
            *    39544    41045: contig of 1502 bp in length
            *    41046    41145: gap of 100 bp
            *    41146    42858: contig of 1713 bp in length
            *    42859    42958: gap of 100 bp
            *    42959    44585: contig of 1627 bp in length
            *    44586    44685: gap of 100 bp
            *    44686    46869: contig of 2184 bp in length
            *    46870    46969: gap of 100 bp
            *    46970    48860: contig of 1891 bp in length
            *    48861    48960: gap of 100 bp
            *    48961    51281: contig of 2321 bp in length
            *    51282    51381: gap of 100 bp
            *    51382    53943: contig of 2562 bp in length
            *    53944    54043: gap of 100 bp
            *    54044    56936: contig of 2893 bp in length
            *    56937    57036: gap of 100 bp
            *    57037    59918: contig of 2882 bp in length
            *    59919    60018: gap of 100 bp
            *    60019    74665: contig of 14647 bp in length
            *    74666    74765: gap of 100 bp
            *    74766    76556: contig of 1791 bp in length
            *    76557    76656: gap of 100 bp
            *    76657    78619: contig of 1963 bp in length
            *    78620    78719: gap of 100 bp
            *    78720    80541: contig of 1822 bp in length
            *    80542    80641: gap of 100 bp
            *    80642    83536: contig of 2895 bp in length
            *    83537    83636: gap of 100 bp
            *    83637    87151: contig of 3515 bp in length
            *    87152    87251: gap of 100 bp
            *    87252    89340: contig of 2089 bp in length
            *    89341    89440: gap of 100 bp
            *    89441    91459: contig of 2019 bp in length
            *    91460    91559: gap of 100 bp
            *    91560    93525: contig of 1966 bp in length
            *    93526    93625: gap of 100 bp
            *    93626    95801: contig of 2176 bp in length
            *    95802    95901: gap of 100 bp
            *    95902    98253: contig of 2352 bp in length
            *    98254    98353: gap of 100 bp
```

```
            *    98354   100651: contig of 2298 bp in length
            *   100652   100751: gap of 100 bp
            *   100752   103409: contig of 2658 bp in length
            *   103410   103509: gap of 100 bp
            *   103510   107033: contig of 3524 bp in length
            *   107034   107133: gap of 100 bp
            *   107134   110878: contig of 3745 bp in length
            *   110879   110978: gap of 100 bp
            *   110979   113971: contig of 2993 bp in length
            *   113972   114071: gap of 100 bp
            *   114072   117157: contig of 3086 bp in length
            *   117158   117257: gap of 100 bp
            *   117258   120573: contig of 3316 bp in length
            *   120574   120673: gap of 100 bp
            *   120674   123933: contig of 3260 bp in length
            *   123934   124033: gap of 100 bp
            *   124034   129503: contig of 5470 bp in length
            *   129504   129603: gap of 100 bp
            *   129604   134903: contig of 5300 bp in length
            *   134904   135003: gap of 100 bp
            *   135004   141319: contig of 6316 bp in length

  Query Match             15.8%;  Score 158.2;  DB 2;  Length 179192;
  Best Local Similarity   59.4%;  Pred. No. 5.9e-27;
  Matches  268;  Conservative    0;  Mismatches  183;  Indels    0;  Gaps    0;

Qy         69 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 128
              | ||||||||||||  ||  |||||| |||        ||| | |||||| |||  ||
Db      88631 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCT 88690

Qy        129 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 188
              |||||| | |       | | |   |||||| |||       | | |   |||||| |||
Db      88691 GCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 88750

Qy        189 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 248
              ||  |||| |  ||       | | |   | ||||   |  ||  |   | ||||  ||
Db      88751 GCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCT 88810

Qy        249 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 308
              |||||| ||| |||  ||  |||||| ||| | |  ||  |||||| ||| |||  ||
Db      88811 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTGCCT 88870

Qy        309 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 368
              |||| |  || |||  ||  |||||| ||| |||  ||  |||||| ||| | |  ||
Db      88871 GCCTCCCAGCATGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCAT 88930

Qy        369 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 428
              |||||| ||| |||  ||  |||||| ||| |||  ||  |||||| |||         ||
Db      88931 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCA 88990

Qy        429 GTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCT 488
              | | | |||| ||| |||       | |      || ||||||   || ||   |  | ||
Db      88991 GCATGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTGCCT 89050

Qy        489 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 519
              ||  | ||||  |   | |||| |  |||||
Db      89051 GCCTGCCTGCCTGCCTGCCTGCAGCACTGCT 89081


RESULT 86
AL583887/c
LOCUS       AL583887              220050 bp    DNA     linear   ROD 09-OCT-2001
DEFINITION  Mouse DNA sequence from clone RP23-121M7 on chromosome 15, complete
            sequence.
ACCESSION   AL583887
VERSION     AL583887.9  GI:16030170
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 220050)
  AUTHORS   Phillimore,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 10, 2001 this sequence version replaced gi:14269983.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-121M7 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP23-121M7 The true
            left end of clone RP23-85M21 is at 142216 in this sequence. The
            true right end of clone RP23-10K12 is at 78004 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .220050
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="15"
                     /clone="RP23-121M7"
                     /clone_lib="RPCI-23"
     repeat_region   28. .453
                     /note="RMER10A repeat: matches 12. .426 of consensus"
     misc_feature    304. .316
                     /note="Sequence from overlapping clone RP23-10K12
                     (AL583889).Assembly confirmed by restriction digest."
     repeat_region   864. .893
                     /note="5 copies 6 mer gctggg 93% conserved"
     repeat_region   873. .968
                     /note="2 copies 48 mer 94% conserved"
     repeat_region   896. .979
                     /note="14 copies 6 mer ggctgg 67% conserved"
     repeat_region   1006. .1123
                     /note="B1_MM repeat: matches 71. .146 of consensus"
     repeat_region   1534. .1680
                     /note="B1_MM repeat: matches 1. .146 of consensus"
     repeat_region   1681. .1788
                     /note="9 copies 12 mer 80% conserved"
     repeat_region   1682. .1801
                     /note="3 copies 40 mer 80% conserved"
     repeat_region   1690. .1793
                     /note="4 copies 26 mer 79% conserved"
     repeat_region   1789. .1933
                     /note="29 copies 5 mer agggg 62% conserved"
     repeat_region   1795. .1938
                     /note="3 copies 48 mer 80% conserved"
     repeat_region   1796. .1933
                     /note="69 copies 2 mer gg 60% conserved"
     repeat_region   1802. .1927
                     /note="21 copies 6 mer gagggg 63% conserved"
     repeat_region   1806. .1935
                     /note="5 copies 26 mer 69% conserved"
     repeat_region   1814. .1930
                     /note="3 copies 39 mer 76% conserved"
     repeat_region   1840. .1923
                     /note="21 copies 4 mer gagg 65% conserved"
```

```
repeat_region   1962. .2129
                /note="B3A repeat: matches 1. .192 of consensus"
repeat_region   2130. .2266
                /note="B1_MM repeat: matches 1. .146 of consensus"
gene            complement(2267. .29223)
                /gene="bM121M7.1"
mRNA            complement(join(2267. .2680,7836. .7999,10910. .10996,
                14639. .14782,15687. .15795,16519. .16653,18289. .18469,
                20448. .20656,23617. .23733,23850. .24010,24647. .24694,
                29101. .29223))
                /gene="bM121M7.1"
                /product="bM121M7.1 (novel protein (ortholog of human
                tubulin tyrosine ligase-like 1 (TTLL1)), isoform 1)"
                /note="match: cDNAs: Em:BC009058 Em:AF104927 Em:AL096883
                Em:BC010510 Em:AL136687 Em:AF173935 Em:AL096886
                match: ESTs: Em:BI359712 Em:BG895852 Em:BF567059
                Em:AW657089 Em:AW901129 Em:BE651256 Em:BI181718
                Em:BI084551"
                /evidence=not_experimental
polyA_site      complement(2267)
                /gene="bM121M7.1"
misc_feature    complement(2283. .2632)
                /gene="bM121M7.1"
                /note="match: GSS: Em:AZ777040"
repeat_region   2304. .2371
                /note="17 copies 4 mer gtgt 80% conserved"
repeat_region   2305. .2382
                /note="3 copies 26 mer 79% conserved"
repeat_region   2311. .2370
                /note="5 copies 12 mer 88% conserved"
polyA_signal    complement(2426. .2431)
                /gene="bM121M7.1"
CDS             complement(join(2551. .2680,7836. .7999,14639. .14782,
                15687. .15795,16519. .16653,18289. .18469,20448. .20656,
                23617. .23729))
                /gene="bM121M7.1"
                /note="match: proteins: Tr:Q9BR27"
                /codon_start=1
                /evidence=not_experimental
                /product="bM121M7.2 (novel protein (ortholog of human
                tubulin tyrosine ligase-like 1 (TTLL1)), isoform 2)"
                /protein_id="CAC93920.1"
                /db_xref="GI:16030171"
                /translation="MAGRVKWVTDIEKSVLINNFEKRGWIQVTENEDWNFYWMSVQTI
                RNVFSVETGYRLSDDQIVNHFPNHYELTRKDLMVKNIKRYRKELEKEGSPLAEKDENG
                KYLYLDFVPVTYMLPADYNLFVEEFRKSPSSTWIMKPCGKAQGKGIFLINKLSQIKKW
                SRDSKTSSFVSQSTKEAYVISVYINNPLLIGGRKFDLRLYVLVSTYRPLRCYMYKLGF
                CRFCTVKYTPSTSELDNMFVHLTNVAIQKHGEDYNHIHGGKWTVNNLRLYLESTRGRE
                VTSKLFDEIHWIIVQSLKAVAVNASPSLTSSTANDRILKYNLINDTLNIAVPNGEIPD
                CKWNKSPPKEVLGNYEILYDEELAQGDGAERELRNRPGQPVGPRAGRSRDSGRSVLTT
                WK"
CDS             complement(join(2551. .2680,7836. .7999,10910. .10996,
                14639. .14782,15687. .15795,16519. .16653,18289. .18469,
                20448. .20656,23617. .23729))
                /gene="bM121M7.1"
                /note="match: proteins: Tr:O95922 Tr:Q9UMU0 Tr:AAH10510
                Tr:AAK92989 Tr:Q9VZ97 Tr:Q9VX74"
                /codon_start=1
                /evidence=not_experimental
                /product="bM121M7.1 (novel protein (ortholog of human
                tubulin tyrosine ligase-like 1 (TTLL1)), isoform 1)"
                /protein_id="CAC93921.1"
                /db_xref="GI:16030172"
                /translation="MAGRVKWVTDIEKSVLINNFEKRGWIQVTENEDWNFYWMSVQTI
                RNVFSVETGYRLSDDQIVNHFPNHYELTRKDLMVKNIKRYRKELEKEGSPLAEKDENG
                KYLYLDFVPVTYMLPADYNLFVEEFRKSPSSTWIMKPCGKAQGKGIFLINKLSQIKKW
                SRDSKTSSFVSQSTKEAYVISVYINNPLLIGGRKFDLRLYVLVSTYRPLRCYMYKLGF
                CRFCTVKYTPSTSELDNMFVHLTNVAIQKHGEDYNHIHGGKWTVNNLRLYLESTRGRE
                VTSKLFDEIHWIIVQSLKAVAPVMNNDKHCFECYGYDIIIDDKLKPWLIEVNASPSLT
                SSTANDRILKYNLINDTLNIAVPNGEIPDCKWNKSPPKEVLGNYEILYDEELAQGDGA
                ERELRNRPGQPVGPRAGRSRDSGRSVLTTWK"
repeat_region   2983. .3138
                /note="B2_Mm2 repeat: matches 1. .157 of consensus"
repeat_region   3147. .3328
                /note="7 copies 26 mer 63% conserved"
repeat_region   3164. .3358
                /note="5 copies 39 mer 69% conserved"
repeat_region   3219. .3302
                /note="3 copies 28 mer 79% conserved"
repeat_region   3256. .3353
                /note="49 copies 2 mer aa 67% conserved"
repeat_region   3269. .3352
                /note="7 copies 12 mer 71% conserved"
misc_feature    complement(3536. .4243)
                /gene="bM121M7.1"
                /note="match: GSS: Em:AQ975102"
repeat_region   3792. .3919
                /note="RSINE1 repeat: matches 4. .131 of consensus"
repeat_region   3920. .4092
                /note="RMER12 repeat: matches 1. .170 of consensus"
repeat_region   4161. .4262
                /note="RMER12 repeat: matches 177. .297 of consensus"
repeat_region   4287. .4575
                /note="RMER17C repeat: matches 114. .414 of consensus"
repeat_region   4576. .5014
                /note="RMER12 repeat: matches 289. .743 of consensus"
repeat_region   5031. .5080
                /note="L1MC4 repeat: matches 6656. .6703 of consensus"
repeat_region   5154. .5183
                /note="5 copies 6 mer gtgtgt 96% conserved"
repeat_region   5239. .5369
                /note="RMER12 repeat: matches 1185. .1315 of consensus"
repeat_region   5422. .5555
                /note="RSINE1 repeat: matches 3. .149 of consensus"
repeat_region   5556. .5682
                /note="B1_MM repeat: matches 1. .125 of consensus"
repeat_region   5713. .5824
                /note="B1_MM repeat: matches 1. .112 of consensus"
misc_feature    5869. .6327
                /note="match: GSS: Em:AZ713136"
repeat_region   6388. .6421
                /note="17 copies 2 mer ac 100% conserved"
repeat_region   6812. .6824
                /note="B4 repeat: matches 74. .86 of consensus"
repeat_region   6825. .6931
```

```
Query Match            15.8%;  Score 158.2;  DB 10;  Length 220050;
Best Local Similarity  59.4%;  Pred. No. 5.8e-27;
Matches  268;  Conservative    0;  Mismatches  183;  Indels    0;  Gaps    0;

Qy         69 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 128
              | ||||||||||||  ||  ||||||  |||        |||  | ||||||  |||  ||
Db      22928 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCT 22869

Qy        129 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 188
              ||||||  | |       | | |   ||||||  |||       | | |   ||||||  |||
Db      22868 GCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 22809

Qy        189 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 248
              ||  ||||  |  ||       | | |   | ||||   |  ||  |   | ||||  ||
Db      22808 GCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCT 22749

Qy        249 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 308
              ||||||  |||  |||  ||  ||||||  |||  | |  ||  ||||||  |||  |||  ||
Db      22748 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCT 22689

Qy        309 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 368
              ||||  |  ||  |||  ||  ||||||  |||  |||  ||  ||||||  |||  |  ||
Db      22688 GCCTCCCAGCATGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCAT 22629

Qy        369 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 428
              ||||||  |||  |||  ||  ||||||  |||  |||  ||  ||||||  |||        ||
Db      22628 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCA 22569

Qy        429 GTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCT 488
              | | | ||||  |||  |||      | |     ||  ||||||   ||  ||   |   | ||
```

```
Db      22568 GCATGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTGCCTGCCT 22509

Qy        489 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 519
              ||  | ||||  |   | |||| |  ||||| .
Db      22508 GCCTGCCTGCCTGCCTGCCTGCAGCACTGCT 22478


RESULT 87
AL583889
LOCUS       AL583889               236344 bp    DNA     linear   ROD 05-APR-2002
DEFINITION  Mouse DNA sequence from clone RP23-10K12 on chromosome 15, complete
            sequence.
ACCESSION   AL583889
VERSION     AL583889.8  GI:16944094
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Tromans,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 15, 2001 this sequence version replaced gi:16596742.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-10K12 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .236344
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="15"
                     /clone="RP23-10K12"
                     /clone_lib="RPCI-23"
BASE COUNT    57826 a  55744 c  58916 g  63858 t
ORIGIN

  Query Match           15.8%;  Score 158.2;  DB 10;  Length 236344;
  Best Local Similarity   59.4%;  Pred. No. 5.8e-27;
  Matches  268;  Conservative    0;  Mismatches  183;  Indels    0;  Gaps    0;

Qy         69 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 128
              | ||||||||||||  ||  |||||| |||        ||| | |||||| |||  ||
Db      55077 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCT 55136

Qy        129 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 188
              ||||||  | |      | | |   |||||| |||      | | |   |||||| |||
Db      55137 GCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 55196

Qy        189 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 248
              ||  |||| |  ||      | | |   | ||||   |  ||  |   | ||||  ||

Db      55197 GCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCT 55256

Qy        249 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 308
              |||||| ||| |||  ||  |||||| ||| | |  ||  |||||| ||| |||  ||
Db      55257 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCT 55316

Qy        309 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 368
              |||| |  || |||  ||  |||||| ||| |||  ||  |||||| ||| | |  ||
Db      55317 GCCTCCCAGCATGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCAT 55376

Qy        369 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 428
              |||||| ||| |||  ||  |||||| ||| |||  ||  |||||| |||         ||
Db      55377 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCA 55436

Qy        429 GTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCT 488
              | | | |||| ||| |||      | |     || ||||||   || ||   |  | ||
Db      55437 GCATGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTGCCT 55496

Qy        489 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 519
              ||  | ||||  |   | |||| |  ||||
Db      55497 GCCTGCCTGCCTGCCTGCCTGCAGCACTGCT 55527


RESULT 88
AL606528
LOCUS       AL606528               148066 bp    DNA     linear   ROD 29-JUN-2002
DEFINITION  Mouse DNA sequence from clone RP23-45H23 on chromosome 13, complete
            sequence.
ACCESSION   AL606528
VERSION     AL606528.7  GI:18873507
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 148066)
  AUTHORS   Oliver,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 22, 2002 this sequence version replaced gi:18476769.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-45H23 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: UK Medical Research Council
            Center code: UK-MRC
            Web site: http://mrcseq.har.mrc.ac.uk
            Contact: mouseq@har.mrc.ac.uk
            --------------.
FEATURES             Location/Qualifiers
     source          1. .148066
                     /organism="Mus musculus"
```

```
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="13"
                     /clone="RP23-45H23"
                     /clone_lib="RPCI-23"
BASE COUNT    43903 a  33152 c  32056 g  38955 t
ORIGIN

  Query Match             15.8%;  Score 157.6;  DB 10;  Length 148066;
  Best Local Similarity   70.3%;  Pred. No. 8.2e-27;
  Matches  211;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        116 TGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTA 175
              | |  | | ||| || |||||||||     | | ||  || |||||||||     | | |
Db      52616 TTCAGGTTTGCTTGCTTGCTTGCTTGCCTGCCTGTTCGCTTGCTTGCTTGCTTGCCTGCC 52675

Qy        176 GGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTA 235
               |   ||||||| ||| || ||| || |      |   |   |  |||| ||||||| |
Db      52676 TGTTCGCTTGCTTGCTTGCTTGCCTGTTCGTTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 52735

Qy        236 GGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 295
               | ||| ||| | || ||||||||||| ||| || |||||||||||| ||| || ||||
Db      52736 TGCCTGTTCGTTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 52795

Qy        296 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 355
              |||||||  | | |  |||||||||||| ||| || |||||||||||| ||| || ||||
Db      52796 TGCTTGCCTGTTCGTTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 52855

Qy        356 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 415
              |||||||| ||| ||||| | | ||||| ||| || ||||||| || ||| | || ||||
Db      52856 TGCTTGCTTGCTTGCCTGTTCGTTTGCTTGCTTGCTTGCTTGCCTGTTCGTTTGCTTGCT 52915


RESULT 89
AL645694
LOCUS       AL645694               189236 bp    DNA     linear   HTG 19-JUN-2002
DEFINITION  Mus musculus chromosome 13 clone RP23-193M3, *** SEQUENCING IN
            PROGRESS ***, 6 unordered pieces..
ACCESSION   AL645694
VERSION     AL645694.5  GI:18151524
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 189236)
  AUTHORS   Matthews,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 15, 2002 this sequence version replaced gi:17017921.
            -------------- Genome Center
            Center: UK Medical Research Council
            Center code: UK-MRC
            Web site: http://mrcseq.har.mrc.ac.uk
            Contact: mouseq@har.mrc.ac.uk
            -------------- Project Information
            Center project name: bM193M3
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 187721 bases at least Q40
            Consensus quality: 188054 bases at least Q30
            Consensus quality: 188428 bases at least Q20
            Insert size: 188736; sum-of-contigs
            Insert size: 192888; 4.5% error; agarose-fp
            Quality coverage: 10.92x in Q20 bases; sum-of-contigs Quality
            coverage: 12.08x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    34757: contig of 34757 bp in length
            *    34758    34857: gap of 100 bp
            *    34858    54711: contig of 19854 bp in length
            *    54712    54811: gap of 100 bp
            *    54812   117926: contig of 63115 bp in length
            *   117927   118026: gap of 100 bp
            *   118027   131915: contig of 13889 bp in length
            *   131916   132015: gap of 100 bp
            *   132016   176088: contig of 44073 bp in length
            *   176089   176188: gap of 100 bp
            *   176189   189236: contig of 13048 bp in length.
FEATURES             Location/Qualifiers
     source          1. .189236
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="13"
                     /clone="RP23-193M3"
                     /clone_lib="RPCI-23"
     misc_feature    1. .34757
                     /note="assembly_fragment:03383
                     clone_end:SP6
                     vector_side:left"
     misc_feature    34858. .54711
                     /note="assembly_fragment:03309.0"
     misc_feature    54812. .117926
                     /note="assembly_fragment:01236
                     fragment_chain:1"
     misc_feature    118027. .131915
                     /note="assembly_fragment:02240
                     fragment_chain:1"
     misc_feature    132016. .176088
                     /note="assembly_fragment:01751
                     fragment_chain:1"
     misc_feature    176189. .189236
                     /note="assembly_fragment:01679
                     fragment_chain:1
                     clone_end:T7
                     vector_side:right"
BASE COUNT    58348 a  41930 c  40027 g  48429 t    502 others
ORIGIN

  Query Match             15.7%;  Score 157.2;  DB 2;  Length 189236;
  Best Local Similarity   70.5%;  Pred. No. 1e-26;
  Matches  210;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy        181 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCT 240
              | ||||| ||| || |||||||      |   |   |  |||| ||||||  |   |
Db    154248 GTTTGCTTGCTTGCTTGCTTGCCTGCCTGTTCGCTTGCTTGCTTGCTTGCCTGCCTGTTC 154307

Qy        241 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 300
              ||| ||| || ||||||| || ||| | || ||||||||||||| ||| || |||||||
Db    154308 GCTTGCTTGCTTGCTTGCCTGTTCGTTTGCTTGCTTGCTTGCTTGCTTGCTTGCCT 154367

Qy        301 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 360
              | ||| | || |||||||||||| ||| || |||||||||||| ||| || |||||||||
Db    154368 GTTCGTTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 154427

Qy        361 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 420
              ||  | | |  |||||||||||| ||| || |||||||||||| ||| || ||||||||
Db    154428 GCCTGTTCGTTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 154487

Qy        421 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGA 478
                | | ||  |  ||||||||||||      | | || ||  ||| || |   || |
Db    154488 GCTTGCCTGTTCGTTTGCTTGCTTGCTTGCCTGTTCGTTTGCTTGCTCTGTTTA 154545
```

```
RESULT 90
AC118179/c
LOCUS       AC118179              217604 bp    DNA     linear   HTG 20-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-448L22, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC118179
VERSION     AC118179.6  GI:25137770
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 217604)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 217604)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 217604)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 20, 2002 this sequence version replaced gi:23908319.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GVMT
                Center clone name: CH230-448L22
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 170842 bases at least Q40
                Consensus quality: 172543 bases at least Q30
                Consensus quality: 173598 bases at least Q20
                Estimated insert size: 173300; sum-of-contigs estimation
                Quality coverage: 10x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    78593: contig of 78593 bp in length
            *    78594    78693: gap of unknown length
            *    78694   216223: contig of 137530 bp in length
            *   216224   216323: gap of unknown length
            *   216324   217604: contig of 1281 bp in length.
FEATURES             Location/Qualifiers
     source          1. .217604
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-448L22"
     misc_feature    1. .1116
                     /note="wgs_contig"
     misc_feature    1. .814
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BZ253397"
     misc_feature    2909. .3666
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BZ253398"
BASE COUNT    48982 a  39593 c  39097 g  46956 t  42976 others
ORIGIN

  Query Match             15.7%;  Score 157.2;  DB 2;  Length 217604;
  Best Local Similarity   52.8%;  Pred. No. 1e-26;
  Matches  339;  Conservative    0;  Mismatches  303;  Indels    0;  Gaps    0;
```

```
Qy        117 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 176
              | ||| | |    |  || ||| |   |  ||| |    | ||| ||| |   |  | | |
Db     189087 GGTTGGTTGGCTGGTTGGTTGGTTGGTTGACTGGTTGGCTGGTTGGTAGGTTGGCTGGTT 189028

Qy        177 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 236
              | ||| ||| | | | | ||| |||     |  |   ||   ||||| | | ||| | |
Db     189027 GGCTGGTTGGTAGGTTGGCTGGTTGGCTGGTTGGTAGGTTGGCTGGTTGGTTGGTTGGTA 188968

Qy        237 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 296
              || ||   | | |  || ||| ||| | | | |   | ||| ||| |   | |  || ||
Db     188967 GGTTGGCTGGTTGGTTGGTTGGTTGGTAGATTGGTAGGTTGGTTGATAAGTAGGTTGGTT 188908

Qy        297 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 356
              | ||| | | | |  || || |||| | | | |  | ||| | | | | | |  || ||
Db     188907 GGTTGGTAGATTGGTTGGTTCCTTGATAGGTTGATAGGTTGGTAGGTTGGTTGGTTGATT 188848

Qy        357 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 416
              | ||| | ||| |  || | | ||| | | | |  || ||| ||| | | | |  || ||
Db     188847 GTTTGGTTGCTAGGTTGGTAGGTTGGTAGGTTGTTTGTTTGGTTGGTTGGTAGGTTGGTT 188788

Qy        417 GCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTT 476
              | |  | |   |  || || ||| ||| || |  |   |   |   |  || |  ||
Db     188787 GGTTGGTTGGTTGGTTGGTTGGTTGGTAGGTTGGTTGGTTGGTAAGTTGGTTCGTT 188728

Qy        477 GATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 536
              | |||   || || | | |  || || |  | | |  || |||   || || | | ||
Db     188727 GGTAGGTTGGTTGGTTGGTAGGTTGGTTGGTTGGTACGTTGGTAGGTTGGTTGGTTGGTA 188668

Qy        537 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCT 596
                | |   || || | || || | |   |     | || || |||   || || | || |
Db     188667 CGTTGGTAGGTTGGTTGGTTGGTTGATAGGTTGGTTGATAGGTTGGTTGGTTGGTT 188608

Qy        597 GCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 656
              | |||   || |  | |   | | |   || || | || || | |   || || | || |
Db     188607 GGTAGGTTGGTTGGTTGATAGGTTGGTAGGTTGGTTGGTTGGTTGGTTGGTTGATAGGTT 188548

Qy        657 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTA 716
              | | |   || || | |  || | |   ||  |  || | ||||| ||| | | |||||
Db     188547 GGTTGGTTGGTTGGTAGATTGGTTGCTTGGTAGATTGGTTGCTTGGTTGGTAGGTTGCTT 188488

Qy        717 GTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGG 758
              |   ||  | || | ||  |     | |||  |   |  | | ||
Db     188487 GGTTGGTAGGTTACTTGGTTGGTTGACTGGTTGGTTGTTTGG 188446
```

RESULT 91
AC110123
LOCUS AC110123 223892 bp DNA linear HTG 18-SEP-2002
DEFINITION Rattus norvegicus clone CH230-319K14, *** SEQUENCING IN PROGRESS ***, 4 unordered pieces.
ACCESSION AC110123
VERSION AC110123.4 GI:23101151
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 223892)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 223892)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (10-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 223892)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (18-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Sep 18, 2002 this sequence version replaced gi:21738195.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequening reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GQKX
Center clone name: CH230-319K14
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 162421 bases at least Q40
Consensus quality: 167280 bases at least Q30
Consensus quality: 170311 bases at least Q20
Estimated insert size: 191280; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

```
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3627: contig of 3627 bp in length
            *     3628     3727: gap of unknown length
            *     3728   220749: contig of 217022 bp in length
            *   220750   220849: gap of unknown length
            *   220850   222487: contig of 1638 bp in length
            *   222488   222587: gap of unknown length
            *   222588   223892: contig of 1305 bp in length.
FEATURES             Location/Qualifiers
     source          1. .223892
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-319K14"
     misc_feature    3728. .5113
                     /note="wgs_contig"
     misc_feature    48568. .50773
                     /note="wgs_contig"
     misc_feature    219069. .220749
                     /note="wgs_contig"
BASE COUNT    48513 a  38895 c  39128 g  46102 t  51254 others
ORIGIN

  Query Match             15.7%;  Score 157.2;  DB 2;  Length 223892;
  Best Local Similarity   52.8%;  Pred. No. 9.9e-27;
  Matches  339;  Conservative    0;  Mismatches  303;  Indels    0;  Gaps    0;

Qy        117 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 176
              | ||| | |   |  || ||| |  |  ||| |   | ||| ||| |  |  | | |
Db      25902 GGTTGGTTGGCTGGTTGGTTGGTTGGTTGACTGGTTGGCTGGTTGGTAGGTTGGCTGGTT 25961

Qy        177 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 236
              | ||| ||| | | | | ||| |||    |  |  ||   ||||| | | ||| | |
Db      25962 GGCTGGTTGGTAGGTTGGCTGGTTGGCTGGTTGGTAGGTTGGCTGGTTGGTTGGTTGGTA 26021

Qy        237 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 296
              || ||   | | |  || ||| ||| | | | |   | ||| ||| |   | |  || ||
Db      26022 GGTTGGCTGGTTGGTTGGTTGGTTGGTAGATTGGTAGGTTGGTTGATAAGTAGGTTGGTT 26081

Qy        297 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 356
              | ||| | | | |  || || |||| | | | |   | ||| | | | | | |  || ||
Db      26082 GGTTGGTAGATTGGTTGGTTCCTTGATAGGTTGATAGGTTGGTAGGTTGGTTGGTTGATT 26141

Qy        357 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 416
              | ||| | ||| |  || | | ||| | | | |  || ||| ||| | | | |  || ||
Db      26142 GTTTGGTTGCTAGGTTGGTAGGTTGGTAGGTTGTTTGTTTGGTTGGTTGGTAGGTTGGTT 26201

Qy        417 GCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTT 476
              | |  |  |   |   || || ||| ||| || |  |   |   |      || |  ||
Db      26202 GGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTAGGTTGGTTGGTTGGTAAGTTGGTTCGTT 26261

Qy        477 GATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 536
              | |||   || || | ||  | | |   || || |  | || |||   || || | |
Db      26262 GGTAGGTTGGTTGGTTGGTAGGTTGGTTGGTTGGTACGTTGGTAGGTTGGTTGGTTGGTA 26321

Qy        537 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCT 596
                | |   || || | || || | |   |     | || || |||   || || | || |
Db      26322 CGTTGGTAGGTTGGTTGGTTGGTTGATAGGTTGGTTGATAGGTTGGTTGGTTGGTT 26381

Qy        597 GCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 656
              | |||   || |  | |   | | |   || || | || || | |   || || | || |
Db      26382 GGTAGGTTGGTTGGTTGATAGGTTGGTAGGTTGGTTGGTTGGTTGGTTGATAGGTT 26441

Qy        657 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTA 716
              | | |   || || | |  || || |||  | | |   ||  |  || | ||||| ||| | | |||||
Db      26442 GGTTGGTTGGTTGGTAGATTGGTTGCTTGGTAGATTGGTTGCTTGGTTGGTAGGTTGCTT 26501

Qy        717 GTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGG 758
                |   ||  | || | || |     | ||| |  |  | | ||
Db      26502 GGTTGGTAGGTTACTTGGTTGGTTGACTGGTTGGTTGTTTGG 26543


RESULT 92
AC117951/c
LOCUS       AC117951              262289 bp    DNA     linear   HTG 22-SEP-2002
DEFINITION  Rattus norvegicus clone CH230-68L11, *** SEQUENCING IN PROGRESS
            ***, 5 unordered pieces.
ACCESSION   AC117951
VERSION     AC117951.6  GI:23194552
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 262289)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 262289)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 262289)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 19, 2002 this sequence version replaced gi:21745850.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequening reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/).  As a result, the
            sequence may extend beyond the ends of the clone and there may be
            contigs that consist entirely of whole genome shotgun sequence
            reads.  Both end sequences and whole genome shotgun sequence only
            contigs will be indicated in the feature table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GTOP
                Center clone name: CH230-68L11
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 247058 bases at least Q40
                Consensus quality: 248642 bases at least Q30
                Consensus quality: 249755 bases at least Q20
                Estimated insert size: 277580; sum-of-contigs estimation
                Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   245055: contig of 245055 bp in length
            *  245056   245155: gap of unknown length
            *  245156   257363: contig of 12208 bp in length
            *  257364   257463: gap of unknown length
            *  257464   258494: contig of 1031 bp in length
            *  258495   258594: gap of unknown length
            *  258595   259652: contig of 1058 bp in length
            *  259653   259752: gap of unknown length
            *  259753   262289: contig of 2537 bp in length.
FEATURES             Location/Qualifiers
     source          1. .262289
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-68L11"
     misc_feature    1. .1619
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    4372. .5390
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    8354. .8807
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH353262"
     misc_feature    complement(242010. .242869)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH353234"
BASE COUNT    68309 a  53803 c  55598 g  72800 t  11779 others
ORIGIN

  Query Match             15.7%;  Score 157.2;  DB 2;  Length 262289;
  Best Local Similarity   55.0%;  Pred. No. 9.8e-27;
  Matches  371;  Conservative    0;  Mismatches  298;  Indels    5;  Gaps    3;

Qy         44 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 103
Db     107825 TGCCTGCCTGCCTGCCTTCCTGCCTTCCTTCTTGCCTGCCTGCCTCCCTGCCTGCCTACC 107766

Qy        104 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 163
Db     107765 TTCCTGCCTGCCTACCTCCCTGCCTGCCTGCCTCCCTACTTGCCTGCCTGCCTGCCTCCC 107706

Qy        164 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 223
Db     107705 TGCCTGCCTCCCTGCCTGCCTGCTGTCT-GCCTGCCTGCCTGCCTGCCTCCCTGCCTGCC 107647

Qy        224 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 283
Db     107646 TGCCTGCCTGTCTGCCTGCCTCCCTCCCTGCCTGCCTGCCTGCCTGCCTCCCTGCCTCCC 107587

Qy        284 CGCTGGCCTGCTTGCTTGCTCGCTG---GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 340
Db     107586 TGCCTGCCTCCCTGCCTGCCTGCTGTCTGCCTGCCTGCCTGCCTGCCTCCCTGCCTGCCT 107527

Qy        341 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
Db     107526 GCCTCCCTGCCTGCCTGCTGTCT-GCCTGCCTGCCTGCCTGCCTGTCTCCCTGCCTGCCT 107468

Qy        401 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTAC 460
Db     107467 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCTGCCTGCCT 107408

Qy        461 GCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 520
Db     107407 CCCTGCCTGCCTGCCTCCCTGCCTGCCTGCCTGCCTCCCTGCCTGCCTCCCTGCCTGCCT 107348

Qy        521 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTA 580
Db     107347 GCCTCCCTGCCTGCCTCCCTGCCTGCCTGCCTACCTCCCTGCCTGCCTGCCTCCCTGCTT 107288

Qy        581 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 640
Db     107287 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCTGCCTGCCTCCCT 107228

Qy        641 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTT 700
Db     107227 GCCTGCCTCCCTGCCTCCCTGCCTCCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 107168

Qy        701 GCTTGCTTGCTTGC 714
Db     107167 GCCTGCCTGCCTGC 107154


RESULT 93
AL928917
LOCUS       AL928917             176961 bp    DNA     linear   VRT 22-JAN-2003
DEFINITION  Zebrafish DNA sequence from clone CH211-103D12, complete sequence.
ACCESSION   AL928917
VERSION     AL928917.8  GI:27848017
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
```

```
REFERENCE   1  (bases 1 to 176961)
  AUTHORS   Dunn,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 22, 2003 this sequence version replaced gi:26984886.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Repeat names
            beginning 'Dr' were identified by the Recon repeat discovery system
            (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr'
            were identified by Rick Waterman (Stephen Johnson lab, WashU).  For
            further information see http://www/Projects/D_rerio/fishmask.shtml
            CH211-103D12 is from a CHORI-211 BAC library
            VECTOR: pTARBAC2.1.
FEATURES             Location/Qualifiers
     source          1. .176961
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-103D12"
                     /clone_lib="CHORI-211"
BASE COUNT    55615 a  33200 c  33486 g  54660 t
ORIGIN

  Query Match           15.7%;  Score 156.8;  DB 5;  Length 176961;
  Best Local Similarity  51.4%;  Pred. No. 1.3e-26;
  Matches  362;  Conservative    0;  Mismatches  342;  Indels    0;  Gaps    0;

Qy         44 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 103
              ||  ||||||   |       |   || |   ||  ||||  || |  || |||||  |
Db     112845 TGGTTGCTTGGTTGGTTAGTTGCTTGGTTTGTTGTTTGCTTACTTGGTTGTTTGCTTGGT 112904

Qy        104 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 163
                |   | |||| ||| | || || ||| | ||  |    |       || |||||
Db     112905 TAGTTGGTTGGATTCTTGGTCGGTTGCTTGGTTGGTTAGTTGCTTGGTTTGTTGTTTGCT 112964

Qy        164 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 223
              |   |   ||  || ||  |  | | | |  | |||| |  |      |   || || |
Db     112965 TACTTGGTTGTTTGCTTGGGTAGTTGGTTGGATTCTTGGTTGGTTAGTTGCTTGGTTGGT 113024

Qy        224 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 283
             ||||| |    | || | |||  | || ||||||||| | |    || |||  | ||
Db     113025 TAGTTGCTTGGTTTGTTGGTTGCTTACTTGGTTGCTTGCTTGGTTAGTTGGTTGGATTCT 113084

Qy        284 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 343
              | | |  |||||| ||| |    | || || ||  ||| | |||  | || ||| |||||
Db     113085 TGGTTGGTTGCTTGGTTGGTTAGTTGCTTGGTTTGTTGTTTGCTTACTTGGTTGTTTGCT 113144

Qy        344 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403
                |||||||||| | | | |  || ||  | | | |  ||  || ||| |
Db     113145 TGGTTAGTTGCTTGCTTGGTTGGTTGGTTTGTTAGTTAATTGGTTGGTTGGGTGGTTGGT 113204

Qy        404 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCC 463
              | | || |  ||| |  | |   |   | |  ||| ||| | ||  ||  |   |
Db     113205 TGATTGCTTATTTGGTTGGTTGGGTGATTGCTTATTTGGTTGGTTAGTTGATTGGTTGAT 113264

Qy        464 TGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 523
              ||  ||   | |||||     || || | || || | |       || | |  | || |
Db     113265 TGGTTGGATGGTTGATTTATTGGTTGGTTGGGTGGTTGATTCATTGGTTGCTTACTTGGT 113324

Qy        524 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTA 583
              |  | || || || | .|    | | || || |||| |      |  | || |  | |
Db     113325 TGATTTCTTGGTTGGTTGCTTAGTTACTTGGTTGCTTGGTTTCTTGGTTGGATAGTTGCT 113384

Qy        584 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 643
              || || | |  |||| |   ||    | || || ||    ||  | |  | || | |
Db     113385 TGGTTGGTTGCTTGCTTGCTTGGTAAGTTGGTTGGTAAGTTGGTGGGTTAGTTGGTTGGT 113444

Qy        644 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCT 703
              | |||| || || | |   || ||   |  || |     |  ||| | || | ||| |
Db     113445 TAGTTGCTTGGTTGGTTGGTTGGTTGGCTGATTGGTTTGTTGTTTGCTTACTTGGTTGTT 113504

Qy        704 TGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCT 747
              |||||| ||| | |    |  ||||| || |   | || || |
Db     113505 TGCTTGGTTGTTTGCTTAGTTGCTTGGTTGGTTGCATGGTTGGT 113548


RESULT 94
AC142424/c
LOCUS       AC142424              140898 bp    DNA     linear   HTG 01-APR-2003
DEFINITION  Rattus norvegicus clone CH230-408K24, WORKING DRAFT SEQUENCE, 39
            unordered pieces.
ACCESSION   AC142424
VERSION     AC142424.2  GI:29423803
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 140898)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A.,
            Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
            Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
            Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G.,
```

```
            Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
            Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
            Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
            Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
            Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
            Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
            Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
            Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
            Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
            Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
            Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
            Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
            Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
            Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
            Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
            Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
            Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 140898)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 140898)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Apr 1, 2003 this sequence version replaced gi:29366848.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KEPT
                Center clone name: CH230-408K24
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 126303 bases at least Q40
                Consensus quality: 130569 bases at least Q30
                Consensus quality: 134257 bases at least Q20
                Estimated insert size: 128743; sum-of-contigs estimation
                Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 39 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1198: contig of 1198 bp in length
            *     1199     1298: gap of unknown length
            *     1299     2537: contig of 1239 bp in length
            *     2538     2637: gap of unknown length
            *     2638     3734: contig of 1097 bp in length
            *     3735     3834: gap of unknown length
            *     3835     5472: contig of 1638 bp in length
            *     5473     5572: gap of unknown length
            *     5573     7555: contig of 1983 bp in length
            *     7556     7655: gap of unknown length
            *     7656     8798: contig of 1143 bp in length
            *     8799     8898: gap of unknown length
            *     8899    10417: contig of 1519 bp in length
            *    10418    10517: gap of unknown length
            *    10518    11625: contig of 1108 bp in length
            *    11626    11725: gap of unknown length
            *    11726    12855: contig of 1130 bp in length
            *    12856    12955: gap of unknown length
            *    12956    14751: contig of 1796 bp in length
            *    14752    14851: gap of unknown length
            *    14852    16352: contig of 1501 bp in length
            *    16353    16452: gap of unknown length
            *    16453    19169: contig of 2717 bp in length
            *    19170    19269: gap of unknown length
            *    19270    21464: contig of 2195 bp in length
            *    21465    21564: gap of unknown length
            *    21565    22701: contig of 1137 bp in length
            *    22702    22801: gap of unknown length
            *    22802    24203: contig of 1402 bp in length
            *    24204    24303: gap of unknown length
            *    24304    26258: contig of 1955 bp in length
            *    26259    26358: gap of unknown length
            *    26359    28665: contig of 2307 bp in length
            *    28666    28765: gap of unknown length
            *    28766    30207: contig of 1442 bp in length
            *    30208    30307: gap of unknown length
            *    30308    32693: contig of 2386 bp in length
            *    32694    32793: gap of unknown length
            *    32794    34145: contig of 1352 bp in length
            *    34146    34245: gap of unknown length
            *    34246    35342: contig of 1097 bp in length
            *    35343    35442: gap of unknown length
            *    35443    37116: contig of 1674 bp in length
            *    37117    37216: gap of unknown length
            *    37217    38848: contig of 1632 bp in length
            *    38849    38948: gap of unknown length
            *    38949    42968: contig of 4020 bp in length
            *    42969    43068: gap of unknown length
            *    43069    46012: contig of 2944 bp in length
            *    46013    46112: gap of unknown length
            *    46113    51041: contig of 4929 bp in length
            *    51042    51141: gap of unknown length
            *    51142    55565: contig of 4424 bp in length
            *    55566    55665: gap of unknown length
            *    55666    59317: contig of 3652 bp in length
            *    59318    59417: gap of unknown length
            *    59418    63593: contig of 4176 bp in length
            *    63594    63693: gap of unknown length
            *    63694    67952: contig of 4259 bp in length
            *    67953    68052: gap of unknown length
            *    68053    73303: contig of 5251 bp in length
            *    73304    73403: gap of unknown length
            *    73404    78136: contig of 4733 bp in length
            *    78137    78236: gap of unknown length
            *    78237    82228: contig of 3992 bp in length
            *    82229    82328: gap of unknown length
            *    82329    88954: contig of 6626 bp in length
            *    88955    89054: gap of unknown length
            *    89055    96366: contig of 7312 bp in length
            *    96367    96466: gap of unknown length
            *    96467   103174: contig of 6708 bp in length
            *   103175   103274: gap of unknown length
            *   103275   113341: contig of 10067 bp in length
            *   113342   113441: gap of unknown length
            *   113442   126269: contig of 12828 bp in length
            *   126270   126369: gap of unknown length
            *   126370   140898: contig of 14529 bp in length.
FEATURES             Location/Qualifiers
     source          1. .140898
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-408K24"
BASE COUNT    37480 a  28929 c  29223 g  41432 t   3834 others
ORIGIN
```

```
  Query Match             15.6%;  Score 155.8;  DB 2;  Length 140898;
  Best Local Similarity   46.3%;  Pred. No. 2.2e-26;
  Matches  348;  Conservative    0;  Mismatches  402;  Indels    1;  Gaps    1;

Qy          7 GAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGT 66
              | ||| |
Db      46119 GCAAATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 46060

Qy         67 ACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGC 126
                                                              ||  ||   |
Db      46059 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTTGGCTGGCTGG 46000

Qy        127 TGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGC 186
              | | |||  ||    |  | | |   | |||  | | ||   ||   || |||   ||  |
Db      45999 TTGGCTGGCTGACTGGCTGGCTGGCTGGCTGGTGGTTGGTTGGTTGGCTGTCTGCC 45940

Qy        187 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGC 246
              | | ||||   ||  |||   ||   || || |  ||| || ||| |  ||   | | |
Db      45939 TGGTTGGCTGTCTGTCTGGTTGGTTTGTTGGTCGACTGCCTGGCTGGCTGGTTGGTTTGT 45880

Qy        247 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 306
              |||   ||| ||| ||| | ||||  |||  || | | | |||    ||   |  || ||
Db      45879 TGGTTGGCTGGCTGGCTGGTTGGCTGGCTGACTGGTTGGTTGGTTGACTGATTGTCTAGC 45820

Qy        307 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 366
              || |  | || || ||| ||||||  | | | |   | | |||   | | ||| |||  |
Db      45819 TGACTGGTTTTCTGGCTGGCTGGCTGGTTGGTTGATTTGTTGGTTGGTTGGCTGGCTGCC 45760

Qy        367 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 426
              |||   ||| ||| ||| | |||  || | | | ||| |||||  ||| ||| |  |
Db      45759 TGGTTGGCTGGCTGGCTGGTTGGTTTGTTGGTTGGCTGGCTGG-CTGGTTGGTTTGTTAG 45701

Qy        427 CAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGG 486
                |||||  ||| ||| |     | | |   | ||| ||| | | ||| | |   ||
Db      45700 TTGGCTGGCTGGTTGGTTGGTTGATTGGCTGGCTGGCTGGCTGGTTGGTTGGTTGACTGG 45641

Qy        487 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 546
              ||    |||   |    |||   |||| | ||  |||||   || || | |   |
Db      45640 TTGTCTAACTGGCTGGTTTTCTGGCTAGCTGGTTGTCTGGCTGGCTGGTTGATTGGTTGA 45581

Qy        547 CTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 606
              ||   ||||| | |   || |  | |||||   |   |||||   |||||   |   ||
Db      45580 TTGGCTGGCTGGTTGATTGGTTGATTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGATTGG 45521

Qy        607 CTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 666
              |  | || || | |   || || | || | | |   || || | || ||   |   ||
Db      45520 TTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGG 45461

Qy        667 CTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTAGTGGGGCCGC 726
              |||   ||||| | |   ||||| |||     ||| ||| ||| ||
Db      45460 CTGACTGGCTGGTGGGCTGGCTGACTGGCTCTTTGGTTGGTTGGTTTTGTTGTCTTTGCT 45401

Qy        727 TTGCCTGCTACTAGGGCTGCTGTGCAGCTGG 757
              | ||||||| |   | ||||     ||| |
Db      45400 ATTCCTGCTATTGCTGTTGCTTATTTGCTTG 45370


RESULT 95
AC128150
LOCUS       AC128150               174032 bp    DNA     linear   HTG 20-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-386E19, *** SEQUENCING IN PROGRESS
            ***.
ACCESSION   AC128150
VERSION     AC128150.3  GI:25139403
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 174032)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174032)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 174032)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 20, 2002 this sequence version replaced gi:23803773.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
```

```
          table.
          -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
              Center project name: GZOV
              Center clone name: CH230-386E19
          -------------- Summary Statistics
              Assembly program: Phrap; version 0.990329
              Consensus quality: 159301 bases at least Q40
              Consensus quality: 160205 bases at least Q30
              Consensus quality: 160799 bases at least Q20
              Estimated insert size: 165067; sum-of-contigs estimation
              Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 1 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *       1   174032: contig of 174032 bp in length.
FEATURES             Location/Qualifiers
     source          1. .174032
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-386E19"
     misc_feature    complement(163561. .164704)
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:RXBAE34TJB"
     misc_feature    complement(163589. .164704)
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:RXBAE34TJC"
     misc_feature    165382. .167144
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    167195. .168910
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    172327. .174032
                     /note="wgs_end_extension
                     clone_end:T7"
BASE COUNT    44840 a  32463 c  33151 g  50641 t  12937 others
ORIGIN

  Query Match             15.4%;  Score 154.4;  DB 2;  Length 174032;
  Best Local Similarity   59.6%;  Pred. No. 4.6e-26;
  Matches  280;  Conservative    0;  Mismatches  186;  Indels    4;  Gaps    1;

Qy         72 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 131
              ||  |||  | | |||||  |||| | |    |||  |  ||  |||| | | |||||
Db      25636 TGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTAGGTTGCTGGAT 25695

Qy        132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191
              |||| | | |  |||  |  ||  |||| | |    | |  |  |   ||||   | |||
Db      25696 TGCTGGATTACTGGATTGCTGGATTGCTGGATTGCTGCATTGCTGCATTGCTGCATTGCT 25755

Qy        192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251
              ||  |||| | |    |||  |  || ||||||| ||||| |       | | ||||| |
Db      25756 GGATTGCTGGATTGCTGGATTGCTGGACTGCTGGATTGCTGGATTACTGGATTGCTGGAC 25815

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
              |||| | ||||| | | ||  | ||||| |   |||||  |||| |  |||| | | ||
Db      25816 TGCTGGATTGCTGGATTGCTGGATTGCTGGACTGCTGGATTGCTGGACTGCTGGATTGCT 25875

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
              | || || | | |||||  |||| |  |||| | | ||  | || || | | |||||
Db      25876 GGATTACTGGATTGCTGGATTGCTGGACTGCTGGATTGCTGGATTACTGGATTGCTGGAC 25935

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTA 431
              |||| | ||||| | | ||  | ||||| | | |||||  |||| |      |||  |
Db      25936 TGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGACTGCTGGATTGCT 25995

Qy        432 GGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCT 491
              || ||||| | |||||    |||| |   |  |||  ||  ||| ||| |  ||  ||||
Db      25996 GGACTGCTGGATTGCT----GGACTGCTGGATTGCTGGACTGCTGGATTGCTGGATTGCT 26051

Qy        492 GGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAG 541
              ||  |||| |   |   | | |   | | |  ||  |||||| |||||||
Db      26052 GGATTGCTGGACTGCTGGATTGCTGGATTGCTGGATTGCTGGACTGCTAG 26101


RESULT 96
AC111479/c
LOCUS       AC111479              253076 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-86B24, *** SEQUENCING IN PROGRESS
            ***, 5 unordered pieces.
ACCESSION   AC111479
VERSION     AC111479.5  GI:30579369
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 253076)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
```

```
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 253076)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 253076)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:24818994.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GMKG
                Center clone name: CH230-86B24
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 238361 bases at least Q40
                Consensus quality: 240714 bases at least Q30
                Consensus quality: 242061 bases at least Q20
                Estimated insert size: 253095; sum-of-contigs estimation
                Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3428: contig of 3428 bp in length
            *     3429     3528: gap of unknown length
            *     3529   247832: contig of 244304 bp in length
            *   247833   247932: gap of unknown length
            *   247933   249082: contig of 1150 bp in length
            *   249083   249182: gap of unknown length
            *   249183   250270: contig of 1088 bp in length
            *   250271   250370: gap of unknown length
            *   250371   253076: contig of 2706 bp in length
FEATURES             Location/Qualifiers
     source          1. .253076
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-86B24"
     misc_feature    1. .1358
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    294. .901
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH308413"
     misc_feature    2061. .3428
                     /note="wgs_contig"
     misc_feature    235405. .236150
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH308371"
     misc_feature    242031. .243421
                     /note="wgs_end_extension
                     clone_end:T7"
BASE COUNT    74614 a  50267 c  48479 g  69285 t  10431 others
ORIGIN

  Query Match             15.4%;  Score 154.4;  DB 2;  Length 253076;
  Best Local Similarity   59.6%;  Pred. No. 4.5e-26;
  Matches  280;  Conservative    0;  Mismatches  186;  Indels    4;  Gaps    1;

Qy         72 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 131
              ||  |||  | | |||||  |||| | |    |||  |  ||  |||| | | |||||
Db      26239 TGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTAGGTTGCTGGAT 26180

Qy        132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191
              |||| | | |  |||  |  ||  |||| | |    | |  |  |  |  ||||   | |||
Db      26179 TGCTGGATTACTGGATTGCTGGATTGCTGGATTGCTGCATTGCTGCATTGCTGCATTGCT 26120

Qy        192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251
              ||  |||| | |    |||  |  || |||||||  |||||  |        | | |||||  |
Db      26119 GGATTGCTGGATTGCTGGATTGCTGGACTGCTGGATTGCTGGATTACTGGATTGCTGGAC 26060

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
              ||||  | |||||  | |  ||  | |||||  |   |||||  ||||  |  ||||  | |  ||
Db      26059 TGCTGGATTGCTGGATTGCTGGATTGCTGGACTGCTGGATTGCTGGACTGCTGGATTGCT 26000

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
              |  ||  ||  |  |  |||||  ||||  |  ||||  |  |  ||  |  ||  ||  |  |  |||||  |
Db      25999 GGATTACTGGATTGCTGGATTGCTGGACTGCTGGATTGCTGGATTACTGGATTGCTGGAC 25940

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTA 431
              ||||  | |||||  | |  ||  | |||||  | |  |||||  ||||  |       |||  |
Db      25939 TGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGACTGCTGGATTGCT 25880

Qy        432 GGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCT 491
              ||  |||||  | |||||    ||||  |   | |||  ||  |||  |||  |  ||  ||||
Db      25879 GGACTGCTGGATTGCT----GGACTGCTGGATTGCTGGACTGCTGGATTGCTGGATTGCT 25824

Qy        492 GGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAG 541
              ||  ||||  |   |  | | |  ||  |||||| |||||||
Db      25823 GGATTGCTGGACTGCTGGATTGCTGGATTGCTGGATTGCTGGACTGCTAG 25774


RESULT 97
AC131433
LOCUS       AC131433              283593 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-88E19, *** SEQUENCING IN PROGRESS
            ***, 7 unordered pieces.
ACCESSION   AC131433
VERSION     AC131433.4  GI:30522205
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
```

```
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 283593)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 283593)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 283593)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24942388.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
               Center project name: GETX
               Center clone name: CH230-88E19
            -------------- Summary Statistics
               Assembly program: Atlas 3.0;
               Consensus quality: 233657 bases at least Q40
               Consensus quality: 237987 bases at least Q30
               Consensus quality: 241136 bases at least Q20
               Estimated insert size: 250075; sum-of-contigs estimation
               Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3892: contig of 3892 bp in length
            *     3893     3992: gap of unknown length
            *     3993   273529: contig of 269537 bp in length
            *   273530   273629: gap of unknown length
            *   273630   274710: contig of 1081 bp in length
            *   274711   274810: gap of unknown length
            *   274811   276051: contig of 1241 bp in length
            *   276052   276151: gap of unknown length
            *   276152   277399: contig of 1248 bp in length
            *   277400   277499: gap of unknown length
            *   277500   278685: contig of 1186 bp in length
            *   278686   278785: gap of unknown length
            *   278786   283593: contig of 4808 bp in length.
FEATURES             Location/Qualifiers
     source          1. .283593
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-88E19"
     misc_feature    3993. .5736
                     /note="wgs_contig"
BASE COUNT    75651 a  53392 c  53293 g  60758 t  40499 others
ORIGIN

  Query Match            15.4%;  Score 154.2;  DB 2;  Length 283593;
  Best Local Similarity  54.1%;  Pred. No. 4.9e-26;
  Matches  335;  Conservative    0;  Mismatches  283;  Indels    1;  Gaps    1;

Qy        140 GAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCT 199
              | || || | | |   ||   || || ||| ||| ||   |  |   | | |  || |
Db     141349 GGGGTGATATTTGCTGTGACAGCAGATGGGGGAGTGGGTAGGTATTTGCTGTTGGTTGGT 141408

Qy        200 TGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCT 259
              || |  |  |   |  || || | | | | |  ||| | | | | |  || |||||
Db     141409 TGGTTGGTTGGTTGGTTGGTTGGTTGGTGGGTGGGTGGGTAGGTGGGTTGGATGGTTGCT 141468

Qy        260 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 319
```

```
Db     141469 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGGA 141528

Qy        320 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 379

Db     141529 TGGTTGGCTGGTTGGTTAGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGATGGT 141588

Qy        380 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 439

Db     141589 TGGTTGGTTGGATCGTTGGTTGGTTGGATGGTTGGTTGGTTGGTTGGTTGCATGGTTGGT 141648

Qy        440 TGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCT 499

Db     141649 TGGTTGGATCATTGGTTGGTTGGTTGGTTGGTTGGTTGGATGGTGGGTTGGTTGGTTGGT 141708

Qy        500 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 559

Db     141709 TGGGTGGTTGTTTGGTTGGATGGTGGGTTGTTTGGTTGGATGGTTGGTTGGTTGGATGGT 141768

Qy        560 AGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCT 619

Db     141769 TGGTTTGTTGGATGGTTGATTGGTTGGATGGTTGGTTGGTTGGATGGTTAGTTGGTTGAA 141828

Qy        620 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 679

Db     141829 TGGTTGCTTGCTTGCTTGCTTGCTTGGATATTTCGTTGGTTGGTTGATTGGTTGGATGGT 141888

Qy        680 AGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTA 739

Db     141889 TGT-TTGTTGGATGTTTAGTTGGTTGAATGGTTGCTTGCTTGCTTGCTTGCTTGGATATT 141947

Qy        740 GGGCTGCTGTGCAGCTGGG 758

Db     141948 TCGTTGGTTGGTTGGTTGG 141966


RESULT 98
AC106435
LOCUS       AC106435               223379 bp    DNA     linear   HTG 08-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-190L22, *** SEQUENCING IN PROGRESS
            ***, 5 unordered pieces.
ACCESSION   AC106435
VERSION     AC106435.3  GI:22857512
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 223379)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 223379)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 223379)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 14, 2002 this sequence version replaced gi:21735213.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GKXE
                Center clone name: CH230-190L22
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 181536 bases at least Q40
                Consensus quality: 186363 bases at least Q30
                Consensus quality: 189368 bases at least Q20
                Estimated insert size: 204062; sum-of-contigs estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
```

```
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   215521: contig of 215521 bp in length
            *  215522   215621: gap of unknown length
            *  215622   217374: contig of 1753 bp in length
            *  217375   217474: gap of unknown length
            *  217475   219103: contig of 1629 bp in length
            *  219104   219203: gap of unknown length
            *  219204   220598: contig of 1395 bp in length
            *  220599   220698: gap of unknown length
            *  220699   223379: contig of 2681 bp in length.
FEATURES             Location/Qualifiers
     source          1. .223379
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-190L22"
     misc_feature    1. .1328
                     /note="wgs_contig"
     misc_feature    197045. .198271
                     /note="wgs_contig"
     misc_feature    202640. .203747
                     /note="wgs_contig"
BASE COUNT    58092 a  36866 c  37682 g  59107 t  31632 others
ORIGIN

  Query Match             15.4%;  Score 153.8;  DB 2;  Length 223379;
  Best Local Similarity   68.6%;  Pred. No. 6.2e-26;
  Matches  212;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy        114 CCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAG 173
              ||||||| || ||| || ||||||||     |   |   ||  | |  ||     |   |
Db     183886 CCTGCTTACTTGCTTGCTTGCTTGCTCGCTCGCTTGCTTGCTCGTTCACTTGCTTGCTTG 183945

Qy        174 TAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAG 233
              ||  ||||||| ||| ||   ||||||     ||  |    |  ||| ||||||| |
Db     183946 ATCGCTCGCTTGCTTGCTCGCTCACTTGCTTGCTCGATCGCTCGCTCGCTCGCTTGCTTG 184005

Qy        234 TAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 293
              |  |||| |||||| ||||||||||||||| ||  ||| ||||||| ||| ||
Db     184006 CTTGCTTGCTGGCTGGCTTGCTTGCTTGCTCGCTTGCTCGCTCGCTTGCTTGCTTGCTGT 184065

Qy        294 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 353
              || ||||||| ||| || |||| ||| ||||||| || ||||| || ||||||| ||  |
Db     184066 CTGGCTTGCTTGCTTGCTTGCTCGCTCGCTCGCTCGCTTGCTTACTCGCTCGCTCGCTCG 184125

Qy        354 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 413
              || ||||||  || ||  |||||||||||||| || |||| ||| ||| ||| || ||
Db     184126 CTCGCTTGCTTACTCGCTCGCTTGCTTGCTCGCTTGCTTGCTCGCTCGCTTGCTTGCTTG 184185

Qy        414 CTTGCTGAG 422
              ||| || ||
Db     184186 CTTTCTAAG 184194


RESULT 99
AC111671
LOCUS       AC111671               265707 bp  DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-50J18, *** SEQUENCING IN PROGRESS
            ***, 6 unordered pieces.
ACCESSION   AC111671
VERSION     AC111671.5  GI:30579213
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 265707)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 265707)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 265707)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:24818905.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
```

```
          genome shotgun sequence reads. Both end sequences and whole genome
          shotgun sequence only contigs will be indicated in the feature
          table.
          -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
              Center project name: GMVD
              Center clone name: CH230-50J18
          -------------- Summary Statistics
              Assembly program: Atlas 3.0;
              Consensus quality: 237575 bases at least Q40
              Consensus quality: 242018 bases at least Q30
              Consensus quality: 245084 bases at least Q20
              Estimated insert size: 248058; sum-of-contigs estimation
              Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 6 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1   258437: contig of 258437 bp in length
          *   258438   258537: gap of unknown length
          *   258538   260145: contig of 1608 bp in length
          *   260146   260245: gap of unknown length
          *   260246   261478: contig of 1233 bp in length
          *   261479   261578: gap of unknown length
          *   261579   263074: contig of 1496 bp in length
          *   263075   263174: gap of unknown length
          *   263175   264393: contig of 1219 bp in length
          *   264394   264493: gap of unknown length
          *   264494   265707: contig of 1214 bp in length.
FEATURES             Location/Qualifiers
     source          1. .265707
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-50J18"
     misc_feature    1627. .2462
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH314712"
     misc_feature    179431. .183168
                     /note="wgs_contig"
     misc_feature    253963. .254736
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH314701"
     misc_feature    255826. .258437
                     /note="wgs_end_extension
                     clone_end:T7"
BASE COUNT    79624 a  48607 c  48049 g  71256 t  18171 others
ORIGIN

  Query Match             15.4%;  Score 153.8;  DB 2;  Length 265707;
  Best Local Similarity   68.6%;  Pred. No. 6.2e-26;
  Matches  212;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy        114 CCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAG 173
              |||||||  ||  |||  ||  ||||||||     |   |    ||  | | | ||      |   |
Db       6104 CCTGCTTACTTGCTTGCTTGCTTGCTCGCTCGCTTGCTTGCTCGTTCACTTGCTTGCTTG 6163

Qy        174 TAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAG 233
                 ||  |||||||  |||  ||   ||||||     ||  |   |   |||  |||||||  |
Db       6164 ATCGCTCGCTTGCTTGCTCGCTCACTTGCTTGCTCGATCGCTCGCTCGCTCGCTTGCTTG 6223

Qy        234 TAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 293
                |  ||||  ||||||  ||||||||||||||||  ||   |||  |||||||  |||  ||
Db       6224 CTTGCTTGCTGGCTGGCTTGCTTGCTTGCTCGCTTGCTCGCTCGCTTGCTTGCTTGCTGT 6283

Qy        294 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 353
              ||  |||||||  |||  ||  ||||  |||  |||||||  ||  |||||  ||  |||||||  ||   |
Db       6284 CTGGCTTGCTTGCTTGCTTGCTCGCTCGCTCGCTCGCTTGCTTACTCGCTCGCTCGCTCG 6343

Qy        354 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 413
              ||  |||||||   ||  ||   ||||||||||||||  ||  ||||  |||  |||  |||  ||  ||
Db       6344 CTCGCTTGCTTACTCGCTCGCTTGCTTGCTCGCTTGCTTGCTCGCTCGCTTGCTTGCTTG 6403

Qy        414 CTTGCTGAG 422
              |||  ||  ||
Db       6404 CTTTCTAAG 6412


RESULT 100
AC095003
LOCUS       AC095003              238046 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-6K12, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC095003
VERSION     AC095003.7  GI:30467553
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 238046)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
```

```
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 238046)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 238046)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 9, 2003 this sequence version replaced gi:24818271.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GBYH
                Center clone name: CH230-6K12
            -------------- Summary Statistics
                Assembly program: Atlas;
                Consensus quality: 226044 bases at least Q40
                Consensus quality: 227924 bases at least Q30
                Consensus quality: 229373 bases at least Q20
                Estimated insert size: 234613; sum-of-contigs estimation
                Quality coverage: 9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   236832: contig of 236832 bp in length
            *   236833   236932: gap of unknown length
            *   236933   238046: contig of 1114 bp in length.
FEATURES            Location/Qualifiers
     source         1. .238046
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-6K12"
     misc_feature   1. .1680
                    /note="wgs_end_extension
                    clone_end:T7"
     misc_feature   6066. .6753
                    /note="clone_boundary
                    clone_end:T7
                    site:EcoRI
                    end_sequence:BH360832"
     misc_feature   228068. .228893
                    /note="clone_boundary
                    clone_end:Sp6
                    site:EcoRI
                    end_sequence:BH360837"
     misc_feature   229310. .230352
                    /note="wgs_end_extension
                    clone_end:Sp6"
     misc_feature   235657. .236832
                    /note="wgs_end_extension
                    clone_end:Sp6"
BASE COUNT    61790 a  49774 c  51497 g  67565 t   7420 others
ORIGIN

  Query Match             15.3%;  Score 152.8;  DB 2;  Length 238046;
  Best Local Similarity   55.5%;  Pred. No. 1.1e-25;
  Matches  315;  Conservative    0;  Mismatches  252;  Indels    1;  Gaps    1;

Qy        150 TAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 209
              | | |||   ||  |   ||   || |||  | | | | |||||   ||| | ||   ||
Db     167418 TTGTCCTTGCTGTCGTTTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGCTGGTTGGCTGG 167477

Qy        210 ACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 269
                |  ||   ||||| | ||| |  ||   | | |||||   | | ||| | | | |||
Db     167478 TTGGCTGGTTGGCTGGTTGGCTGGCTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGG 167537

Qy        270 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 329
              | | | | | | | | ||||  | | | | ||| | |||   ||| ||| | | |||||
Db     167538 CTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGCTGGCTGGTTGGCTGG 167597

Qy        330 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 389
              || ||| ||| | | | |  || ||| ||| | |   |  || ||| |||   |   |
Db     167598 -TTGGTTGGTTGGTTGGTTGGGTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGCTGGCTG 167656

Qy        390 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAA 449
              ||  || ||| | | | | ||| ||| |  |  | | |   || ||  || ||| |
Db     167657 GTTGGCTGGTTGGTTGGTTGGCTGGTTGTTTGGTTGGCTGGCTGGTTGGCTGGTTGGTTG 167716

Qy        450 GGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 509
              |  | | |   |  || ||| | | ||| | |   ||  || | ||||| | |   |||||
Db     167717 GTTGGCTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGCTG 167776

Qy        510 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTC 569
              | || || | |   || ||   || || | |   |||||  | || ||   |   || |
Db     167777 GTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGTTG 167836

Qy        570 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTG 629
              || || | | | || || | || || | |   ||||  | ||||| | |   || ||
Db     167837 GCTGGTTGGTTGGATGGTTGGTTGGTTGGTTGGCTGGTTGGCTGGTTGGTTGGTTG 167896

Qy        630 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 689
              || ||   |   || || | || || | |   |||| | ||||| |||   || ||
Db     167897 GCTGGTTGGCTGGTTGGTTGGTTGGATGGTTGGTTGGCTAGTTGGCTGGTAGGGTGGTTG 167956

Qy        690 CCTGCTGGCTTGCTTGCTTGCTTGCTAG 717
              || | |  || ||| ||| ||| | |
Db     167957 GTTGGTTGACTGGTTGGTTGGTTGGTTG 167984


Search completed: August 21, 2003, 23:29:52
Job time : 3770.64 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 21, 2003, 18:39:57 ; Search time 281.859 Seconds
(without alignments)
9577.262 Million cell updates/sec

Title: US-09-831-000-1_COPY_1_1000
Perfect score: 1000
Sequence: 1 gatcgggaaaacgcgagggg.....gtgcacggccaagttacaag 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1000	100.0	133719	21	AAC64754	Macaca mulatta rha
	2	791.6	79.2	128139	24	AAI64291	RRV genome nucleot
	3	192.8	19.3	576	22	ABA41262	Probe #19728 for g
	4	192.8	19.3	576	22	AAI57480	Probe #26166 used
	5	180.6	18.1	555	22	ABA31139	Probe #9605 for ge
	6	180.6	18.1	555	22	AAI44102	Probe #12788 used
c	7	144.8	14.5	133719	21	AAC64754	Macaca mulatta rha
c	8	134.2	13.4	128139	24	AAI64291	RRV genome nucleot
c	9	127.6	12.8	35100	20	AAV73802	KSHV LUR DNA (nucl
c	10	127.6	12.8	137507	19	AAV19941	KSHV long unique c
	11	103.4	10.3	385	16	AAQ95095	Configuration dete
	12	101.8	10.2	344	16	AAQ84040	Plasmid pCU19. Sy
	13	99.6	10.0	14212	22	AAK79692	Human immune/haema
	14	99.6	10.0	14212	22	AAK80024	Human immune/haema
	15	99.6	10.0	14212	22	AAK85921	Human immune/haema
c	16	99.2	9.9	385	16	AAQ95095	Configuration dete
c	17	96.8	9.7	344	16	AAQ84040	Plasmid pCU19. Sy
c	18	91.6	9.2	2130	25	ABZ39851	N. gonorrhoeae nuc
c	19	90	9.0	2855	22	ABA15389	Human nervous syst
c	20	84.4	8.4	15682	22	AAK89506	Human digestive sy
	21	80.8	8.1	42999	24	ABS65032	Invertebrate forag
	22	80.8	8.1	42999	25	ABX11086	Human ribosomal RN
	23	79.8	8.0	1830121	17	AAT42063	Haemophilus influe
	24	79	7.9	1172	24	ABQ31994	Oligonucleotide fo
c	25	79	7.9	1172	24	ABQ31995	Oligonucleotide fo
	26	79	7.9	1172	24	ABQ31996	Oligonucleotide fo
c	27	79	7.9	1172	24	ABQ31997	Oligonucleotide fo
c	28	78.4	7.8	388	24	ABQ97027	Mouse ES cell rela
c	29	78.4	7.8	428	24	ABQ97101	Mouse ES cell rela
c	30	78.4	7.8	436	24	ABQ97086	Mouse ES cell rela
c	31	78	7.8	169	24	AAL40273	BASB204 related DN
c	32	77.4	7.7	409	24	ABQ97094	Mouse ES cell rela
c	33	77.4	7.7	1172	24	ABQ31994	Oligonucleotide fo
	34	77.4	7.7	1172	24	ABQ31995	Oligonucleotide fo
c	35	77.4	7.7	1172	24	ABQ31996	Oligonucleotide fo
	36	77.4	7.7	1172	24	ABQ31997	Oligonucleotide fo
	37	77.2	7.7	368	22	AAL06216	Human reproductive
	38	77.2	7.7	368	22	AAL07359	Human reproductive
c	39	77.2	7.7	392	24	ABQ97468	Mouse ES cell rela
	40	76.8	7.7	7202	22	ABA19579	Human nervous syst
	41	76.8	7.7	7202	22	AAK71750	Human immune/haema
c	42	75.8	7.6	410	24	ABQ97057	Mouse ES cell rela
	43	75.4	7.5	24512	22	ABA19578	Human nervous syst
	44	75.4	7.5	24512	22	AAK67571	Human immune/haema
	45	75.4	7.5	24512	22	AAK71749	Human immune/haema
c	46	74.4	7.4	427	24	ABQ96982	Mouse ES cell rela
c	47	72.8	7.3	379	24	ABQ96978	Mouse ES cell rela
c	48	72.6	7.3	368	20	AAV88624	EST clone GX576.
	49	71.8	7.2	16084	22	AAK69442	Human immune/haema
c	50	71.8	7.2	16084	22	AAK82881	Human immune/haema
c	51	70.8	7.1	157	24	AAL40272	BASB204 related DN
	52	70.6	7.1	600	6	AAN50853	Sequence encoding
	53	69.6	7.0	3249	23	AAS76504	DNA encoding novel
	54	69.4	6.9	7614	22	AAK85846	Human immune/haema
	55	69.2	6.9	6286	22	AAS46592	Tumour suppressor
c	56	67.2	6.7	3588	23	AAS70903	DNA encoding novel
c	57	67.2	6.7	3588	23	AAS92599	DNA encoding novel
	58	67	6.7	397	24	ABQ97197	Mouse ES cell rela
c	59	66.8	6.7	183	24	AAL40271	BASB204 related DN
c	60	66.8	6.7	186	24	AAL40264	BASB204 related DN
c	61	66.8	6.7	228	24	AAL40265	BASB204 related DN
c	62	66.8	6.7	240	24	AAL40268	BASB204 related DN
c	63	66.8	6.7	252	24	AAL40266	BASB204 related DN
c	64	66.8	6.7	264	24	AAL40267	BASB204 related DN
c	65	66.8	6.7	267	24	AAL40269	BASB204 related DN
c	66	66.8	6.7	267	24	AAL40270	BASB204 related DN
	67	66.2	6.6	13321	22	AAS46422	Tumour suppressor
c	68	66	6.6	147	25	ABZ40780	N. gonorrhoeae nuc
	69	65.8	6.6	223	22	AAL36820	Human musculoskele
	70	65.8	6.6	223	22	AAK64806	Human immune/haema
	71	65.8	6.6	223	25	ABX59808	cDNA encoding nove
	72	65.4	6.5	230	22	AAK54984	Human immune/haema
c	73	65.2	6.5	393	24	ABQ97710	Mouse ES cell rela
	74	65.2	6.5	31718	22	AAS39915	Genomic sequence #
	75	65.2	6.5	31718	22	AAS39916	Genomic sequence #
	76	65.2	6.5	31718	22	AAL06415	Human reproductive
	77	65.2	6.5	31718	22	AAL06416	Human reproductive
c	78	65.2	6.5	31718	22	AAK73104	Human immune/haema
c	79	65.2	6.5	31718	22	AAK73120	Human immune/haema
	80	65.2	6.5	31718	22	AAK87442	Human immune/haema
	81	65.2	6.5	31718	22	AAK87443	Human immune/haema

c	82	65.2	6.5	31718	22	AAK87573
c	83	65.2	6.5	31718	22	AAK87592
	84	65.2	6.5	31718	22	AAK90359
	85	65.2	6.5	31718	22	AAK90360
c	86	65	6.5	148834	24	ABK83570
c	87	64.4	6.4	32249	22	ABA17155
	88	64.2	6.4	223	22	AAL36821
	89	64.2	6.4	223	22	AAK64807
	90	64.2	6.4	223	25	ABX59809
c	91	63.4	6.3	3595	20	AAX08988
c	92	63.4	6.3	3595	25	ABX12095
	93	63.4	6.3	8174	24	AAL44951
c	94	62.8	6.3	149	24	AAL40260
c	95	62.8	6.3	149	24	AAL40261
c	96	62.8	6.3	149	24	AAL40262
c	97	62.8	6.3	149	24	AAL40263
c	98	62.8	6.3	2084	24	AAL40256
c	99	62.8	6.3	2084	24	AAL40257
	100	62.8	6.3	6109	24	ABL32327
	101	62.8	6.3	6109	24	AAS61078
c	102	61.8	6.2	1771	18	AAT62138
c	103	61.8	6.2	1771	19	AAV47558
c	104	61.8	6.2	1771	20	AAZ25621
c	105	61.8	6.2	1771	24	AAF88524
c	106	61.8	6.2	1771	24	AAD40285
c	107	61.8	6.2	1771	24	ABK81733
c	108	61.8	6.2	1771	24	AAS96022
c	109	61.6	6.2	1542	10	AAN90025
c	110	61.2	6.1	10483	22	AAK80686
	111	60.8	6.1	2239	24	ABK90515
c	112	60.8	6.1	45546	20	AAX23520
c	113	60.6	6.1	15720	22	AAI29504
c	114	60.6	6.1	15720	24	ABL62358
c	115	60.6	6.1	15720	25	ABZ33690
	116	60.4	6.0	1951	24	ABQ19256
c	117	60.4	6.0	1951	24	ABQ19257
c	118	60.4	6.0	6171	24	ABL32789
c	119	60.4	6.0	168174	24	ABT11173
c	120	60.4	6.0	168273	24	ABT11114
c	121	60.2	6.0	1412	24	AAL46206
	122	59.4	5.9	15682	22	AAK89506
	123	59.2	5.9	160820	25	ABQ76673
c	124	58.8	5.9	159095	25	ABZ80818
c	125	58.2	5.8	186	21	AAC14822
	126	58.2	5.8	18376	22	AAK67167
	127	57.4	5.7	194	20	AAX01333
c	128	56.8	5.7	350	16	AAQ95194
c	129	56.6	5.7	412	22	AAS38476
c	130	56	5.6	573	22	ABA64355
c	131	56	5.6	573	22	ABA76635
c	132	56	5.6	573	22	ABA31495
c	133	56	5.6	573	22	ABA41143
c	134	56	5.6	573	22	AAK12816
c	135	56	5.6	573	22	AAK25269
c	136	56	5.6	573	22	AAK38543
c	137	56	5.6	573	22	AAK51276
c	138	56	5.6	573	24	ABS12615
c	139	56	5.6	573	24	ABS24815
	140	55.8	5.6	8129	22	AAS46763
c	141	55.8	5.6	113033	25	AAL54213
c	142	55.6	5.6	137	24	AAL40258
c	143	55.6	5.6	137	24	AAL40259
c	144	55.6	5.6	1951	24	ABQ19254
	145	55.6	5.6	1951	24	ABQ19255
c	146	55.6	5.6	32187	22	AAK89946
	147	55.4	5.5	190	20	AAX01332
c	148	55.4	5.5	27048	22	AAS02354
c	149	54.8	5.5	345	24	ABQ96922
c	150	54.8	5.5	12007	24	ABL32716
c	151	54.6	5.5	354	24	ABQ97167
	152	54.6	5.5	98472	24	ABS54549
c	153	54.4	5.4	538	16	AAQ95198
	154	54.2	5.4	1951	24	ABQ19254

c	155	54.2	5.4	1951	24	ABQ19255
	156	54.2	5.4	462586	25	ABQ84281
	157	53.8	5.4	356	23	AAS76990
c	158	53.8	5.4	789	20	AAX40044
c	159	53.8	5.4	2409	22	AAH17441
c	160	53.8	5.4	4065	22	AAD16913
c	161	53.8	5.4	72604	20	AAZ10752
c	162	53.8	5.4	72604	24	ABK43231
	163	53.6	5.4	4182	22	AAH26524
	164	53.6	5.4	26555	22	AAK68372
	165	53.6	5.4	26555	22	AAK68605
	166	53.6	5.4	26555	22	AAI62833
	167	53.6	5.4	114955	20	AAX53491
	168	53.6	5.4	136328	24	ABZ35015
c	169	53.6	5.4	188888	24	ABQ75562
	170	53.4	5.3	7045	24	ABS97907
c	171	53.4	5.3	319608	21	AAH51601
c	172	53.4	5.3	319608	22	AAS09301
c	173	53.2	5.3	26166	25	ACC50869
c	174	53.2	5.3	26166	25	ABZ71491
	175	53.2	5.3	34634	24	AAD31198
c	176	53.2	5.3	97835	24	ABK84796
c	177	53	5.3	1960	18	AAT96724
c	178	53	5.3	3496	20	AAZ27622
c	179	53	5.3	3766	20	AAZ27621
c	180	53	5.3	4936	20	AAZ27618
c	181	52.6	5.3	397658	25	ABQ83210
	182	52.2	5.2	178	20	AAX01330
	183	52.2	5.2	678	24	ABQ28082
c	184	52.2	5.2	678	24	ABQ28083
c	185	52.2	5.2	119198	24	ABK83565
	186	52	5.2	573	22	ABA64355
	187	52	5.2	573	22	ABA76635
	188	52	5.2	573	22	ABA31495
	189	52	5.2	573	22	ABA41143
	190	52	5.2	573	22	AAK12816
	191	52	5.2	573	22	AAK25269
	192	52	5.2	573	22	AAK38543
	193	52	5.2	573	22	AAK51276
	194	52	5.2	573	24	ABS12615
	195	52	5.2	573	24	ABS24815
c	196	51.8	5.2	2408	21	AAA28523
	197	51.6	5.2	182	20	AAX01331
c	198	51.4	5.1	155074	24	ABN85735
	199	51.2	5.1	201143	24	ABK83568
c	200	51	5.1	369	24	ABL82290
c	201	51	5.1	497	21	AAA82163
c	202	50.8	5.1	1559	18	AAT91855
	203	50.8	5.1	659158	25	ABX16390
	204	50.6	5.1	22655	22	AAK70122
c	205	50.4	5.0	1682	22	AAD20196
c	206	50.4	5.0	1682	22	AAF86072
c	207	50.4	5.0	1682	24	ABK83618
c	208	50.2	5.0	2000	24	ABZ17227
c	209	50	5.0	572	24	ABQ47268
	210	50	5.0	572	24	ABQ47269
c	211	50	5.0	10577	21	AAA81494
	212	50	5.0	22400	25	ABT16305
c	213	50	5.0	349980	21	AAF21610
c	214	50	5.0	1437668	21	AAA81490
c	215	49.8	5.0	348	18	AAT65077
c	216	49.8	5.0	348	20	AAX17814
	217	49.8	5.0	6292	22	AAS46735
c	218	49.8	5.0	9831	22	ABA07250
c	219	49.8	5.0	9831	22	AAK89891
	220	49.6	5.0	7597	24	ABL33012
	221	49.6	5.0	90541	24	ABS52847
c	222	49.4	4.9	519	22	ABA61017
c	223	49.4	4.9	519	22	AAK09310
c	224	49.4	4.9	519	22	AAK35199
c	225	49.4	4.9	519	22	AAI40915
c	226	49.4	4.9	519	23	ABS34947
c	227	49.4	4.9	1337	20	AAZ17263

Human immune/haema
Human immune/haema
Human digestive sy
Human digestive sy
Human cDNA differe
Human nervous syst
Human musculoskele
Human immune/haema
cDNA encoding nove
Sequence encoding
Human cerberus-lik
Equine herpesvirus
BASB204 related DN
BASB204 related DN
BASB204 related DN
BASB204 related DN
BASB204 related DN
BASB204 related DN
Human immune syste
Human gene regulat
Leishmania tropica
Leishmania antigen
Leishmania tropica
L. tropica LT-210
Leishmania tropica
Leishmania antigen
Leishmania antigen
DNA encoding antig
Human immune/haema
Human pyroglutamyl
Human kidney amino
C899P determined c
Colon adenocarcino
Human colon tumour
Oligonucleotide fo
Oligonucleotide fo
Human immune syste
Human 5-lipoxygena
Human 5-lipoxygena
Human liver cancer
Human digestive sy
Androgen receptor
Human PAI-2 polymo
Human secreted pro
Human immune/haema
Allelic ladder, HU
Simple tandem repe
Novel human diagno
Human foetal liver
Human foetal liver
Probe #9961 for ge
Probe #19609 for g
Human brain expres
Human brain expres
Human bone marrow
Human bone marrow
Human genome-deriv
Human genome-deriv
Tumour suppressor
SR protein-specifi
BASB204 related DN
BASB204 related DN
Oligonucleotide fo
Oligonucleotide fo
Human digestive sy
Allelic ladder, HU
Porcine Gal alpha(
Mouse ES cell rela
Human immune syste
Mouse ES cell rela
Human multidrug re
Simple tandem repe
Oligonucleotide fo

Oligonucleotide fo
Chromosome 2 Asthm
DNA encoding novel
Prostate cancer as
Human cDNA sequenc
Human phorbol acti
Genomic sequence o
Human HKNG1 gene.
Human proton/oligo
Human immune/haema
Human immune/haema
Human genomic DNA
Human adenosine A1
Human gene express
Human related CYP
Human UDP-glucuron
Human chromosome 1
Human schizophreni
Human secreted pro
Secreted protein g
Human WKL-1 genomi
Human cDNA differe
Brassica sp. tapet
Plasmid TOPI-1. S
Plasmid TOP-1. Sy
Plasmid TOG-1. Sy
Human transporter
Allelic ladder, HU
Oligonucleotide fo
Oligonucleotide fo
Human cDNA differe
Human foetal liver
Human foetal liver
Probe #9961 for ge
Probe #19609 for g
Human brain expres
Human brain expres
Human bone marrow
Human bone marrow
Human genome-deriv
Human genome-deriv
Human opioid growt
Allelic ladder, HU
Human genomic regi
Human DNA differen
Human ovarian canc
N. meningitidis pa
DUB-1 enhancer/pro
Mouse high growth
Human immune/haema
Human translation
Human translation
Human cDNA differe
Arabidopsis thalia
Oligonucleotide fo
Oligonucleotide fo
N. meningitidis pa
Zinc finger protei
Neisseria meningit
N. meningitidis B
Canine genomic mic
Microsatellite rep
Tumour suppressor
Human pancreatic c
Human digestive sy
Human immune syste
Human SR protein-s
Human foetal liver
Human brain expres
Human bone marrow
Probe #9601 used t
Human liver single
Human gene express

	228	49.4	4.9	3593	22	AAD02774	Murine B7-4 protei
	229	49.4	4.9	3593	22	AAD02709	Murine B7-4 protei
	230	49.4	4.9	3593	25	AAD51099	Murine B7-4 DNA.
	231	49.4	4.9	3593	25	ABT14004	Murine PD-L1 codin
	232	49.4	4.9	12237	24	ABL34359	Human immune syste
c	233	49.2	4.9	3092	23	AAS69141	DNA encoding novel
c	234	49.2	4.9	10732	21	AAA10594	Gene encoding a su
	235	49.2	4.9	72604	20	AAZ10752	Genomic sequence o
	236	49.2	4.9	72604	24	ABK43231	Human HKNG1 gene.
	237	49.2	4.9	249487	24	ABN85733	Mouse genomic regi
c	238	49	4.9	2348	21	AAA28525	Human opioid growt
	239	49	4.9	157875	24	ABK99972	Human CADPKL genom
c	240	48.8	4.9	650	16	AAT01589	Rat neu promoter.
	241	48.8	4.9	28068	21	AAA59650	Modified adenovira
	242	48.8	4.9	30115	21	AAA59651	Modified adenovira
	243	48.6	4.9	174	20	AAX01329	Allelic ladder, HU
	244	48.4	4.8	120	16	AAQ84038	Plasmid pCU19 olig
c	245	48.4	4.8	120	16	AAQ84038	Plasmid pCU19 olig
	246	48.4	4.8	120	16	AAQ95093	Configuration dete
c	247	48.4	4.8	120	16	AAQ95093	Configuration dete
	248	48.4	4.8	405	20	AAX89576	Autonomously repli
	249	48.4	4.8	405	21	AAL44929	Multiple tandem in
c	250	48.4	4.8	836	25	ABZ75079	Mouse embryonic st
c	251	48.4	4.8	1623	25	ABZ75087	Mouse embryonic st
c	252	48.4	4.8	149671	24	ABK84797	Human cDNA differe
	253	48.2	4.8	170	20	AAX01328	Allelic ladder, HU
c	254	48.2	4.8	381	23	AAS74660	DNA encoding novel
	255	48.2	4.8	57728	22	AAC87588	Human 9p11 chromos
	256	48	4.8	1607	21	AAC59256	Human secreted pro
	257	48	4.8	7814	22	AAS46529	Tumour suppressor
	258	48	4.8	2155561	24	ABN71527	Streptococcus poly
	259	47.8	4.8	3228	22	AAC90434	Murine Lymph node
	260	47.8	4.8	7311	22	AAD07025	Pseudomonas aerugi
c	261	47.6	4.8	1951	24	ABQ19256	Oligonucleotide fo
	262	47.6	4.8	1951	24	ABQ19257	Oligonucleotide fo
c	263	47.6	4.8	8482	23	ABL13462	Drosophila melanog
	264	47.6	4.8	16200	24	AAD36438	Mouse L66 genomic
c	265	47.4	4.7	1891	24	ABN69176	Streptococcus poly
c	266	47.4	4.7	2289	21	AAA28526	Human opioid growt
c	267	47.4	4.7	2290	21	AAA28522	Human opioid growt
	268	47.4	4.7	6419	24	ABL32266	Human immune syste
	269	47.4	4.7	103747	24	ABQ88139	Human osteoblast d
c	270	47.2	4.7	1893	24	ABN70522	Streptococcus poly
	271	47	4.7	502	24	ABQ92814	Triticum tauschii/
	272	47	4.7	1412	24	AAL46206	Human liver cancer
	273	47	4.7	10301	22	AAK84949	Human immune/haema
	274	47	4.7	13304	25	AAD51620	Human cytochrome P
	275	47	4.7	14776	24	ABK83747	Human cDNA differe
	276	47	4.7	14776	24	ABN95590	Gene #2088 used to
c	277	47	4.7	18443	20	AAV83942	Bacterial artifici
	278	47	4.7	160755	23	AAH88704	Human DNA sequence
	279	47	4.7	1691080	24	ABX08336	Human phosphodiest
c	280	46.8	4.7	705	24	ABQ36954	Oligonucleotide fo
	281	46.8	4.7	705	24	ABQ36955	Oligonucleotide fo
	282	46.8	4.7	8622	22	AAF58432	Murine mOCIL genom
	283	46.8	4.7	38186	20	AAZ32028	Human METH1 relate
	284	46.8	4.7	38186	22	AAC90085	AC004449 cDNA clon
c	285	46.8	4.7	129722	24	ABQ88117	Human osteoblast d
c	286	46.6	4.7	275	16	AAQ85534	DNA probe 40 detec
	287	46.6	4.7	1527	23	AAS86382	DNA encoding novel
	288	46.6	4.7	2782	23	AAS83693	DNA encoding novel
c	289	46.6	4.7	86574	24	ABK83560	Human cDNA differe
c	290	46.4	4.6	124	16	AAQ95211	Simple tandem repe
c	291	46.4	4.6	124	16	AAQ84037	Plasmid pCU19 olig
c	292	46.4	4.6	124	16	AAQ95092	Configuration dete
	293	46.4	4.6	4603	22	AAK78076	Human immune/haema
c	294	46.4	4.6	11729	24	ABL32894	Human immune syste
	295	46.2	4.6	124	16	AAQ84037	Plasmid pCU19 olig
	296	46.2	4.6	124	16	AAQ95092	Configuration dete
c	297	46.2	4.6	693	22	AAL36056	Human musculoskele
c	298	46.2	4.6	693	25	ABX59044	cDNA encoding nove
	299	46.2	4.6	11084	20	AAX22302	Human IL-1ra BAC c
c	300	46.2	4.6	11901	20	AAX02998	Human IL-1ra BAC c

ALIGNMENTS

RESULT 1
AAC64754
ID AAC64754 standard; DNA; 133719 BP.
XX
AC AAC64754;
XX
DT 28-FEB-2001 (first entry)
XX
DE Macaca mulatta rhadinovirus 17577 (RRV) genome sequence SEQ ID NO:1.
XX
KW Macaca mulatta rhadinovirus 17577; RRV; rhesus macaque rhadinovirus;
KW genome; Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6;
KW IL-6; macrophage inflammatory protein; MIP; diagnosis; vaccine;
KW cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma;
KW lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy;
KW splenomegaly; hypergammaglobinulinaemia; autoimmune haemolytic anaemia;
KW ds.
XX
OS Macaca mulatta rhadinovirus 17577.
XX
PN WO200028040-A2.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26260.
XX
PR 06-NOV-1998; 98US-0107507.
PR 20-NOV-1998; 98US-0109409.
XX
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI Wong SW, Axthelm MK, Searles RP;
XX
DR WPI; 2000-376552/32.
XX
PT New rhesus rhadino virus for producing non-human primate model useful
PT for testing potential treatments and efficacy of the candidate vaccine
PT for conditions associated with RRV infection -
XX
PS Claim 2; Page 83-122; 141pp; English.
XX
CC The present invention describes a novel rhesus macaque rhadinovirus
CC called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the
CC RRV genome sequence, and AAB53123 to AAB53204 represent the proteins
CC encoded by the genome sequence. The present invention also specifically
CC claims the individual open reading frame (ORF) nucleotide sequences from
CC the genome which encode the individual proteins, but these sequences are
CC not given. A non-human animal infected with RRV can be used for testing
CC the efficacy of drug in the treatment of condition associated with
CC infection with RRV such as Kaposi's sarcoma, lymphoproliferative
CC disorders, B-cell hyperplasia, lymphadenopathy, splenomegaly,
CC hypergammaglobinulinaemia or autoimmune haemolytic anaemia, by
CC administering the drug to a immuno-compromised non-human primate
CC preferably Rhesus macaque monkey obtained by as a result of infection
CC by Simian Immunodeficiency Virus (SIV). RRV is useful for producing
CC non-human primate model for testing potential treatments for conditions
CC associated with RRV infection. It is also useful for testing the
CC efficacy of the candidate vaccine against RRV infection or conditions
CC associated with its infection by administering the vaccine to the
CC subject capable of infection with RRV, inoculating the subject with RRV
CC and observing the effect of vaccine. AAC64755 to AAC64765 and AAB53205
CC to AAB53213 represent sequence used in the exemplification of the
CC present invention.
XX
SQ Sequence 133719 BP; 32746 A; 35648 C; 34521 G; 30804 T; 0 other;

Query Match 100.0%; Score 1000; DB 21; Length 133719;
Best Local Similarity 100.0%; Pred. No. 3.3e-245;

```
  Matches 1000;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACA 60

Qy         61 CCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 120

Qy        121 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCT 180

Qy        181 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCT 240

Qy        241 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 300

Qy        301 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 360

Qy        361 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 420

Qy        421 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 480

Qy        481 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 540

Qy        541 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 600

Qy        601 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 660

Qy        661 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGG 720

Qy        721 GGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGGGCTGCCGGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGGGCTGCCGGC 780

Qy        781 CAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACCCGGGCCTCTCCCCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACCCGGGCCTCTCCCCC 840

Qy        841 AGGCACAAAGCAGTAGGGCTGGCCAGGGGAAACAGTAAGGCTGATTGCTTGCTGAAAAAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AGGCACAAAGCAGTAGGGCTGGCCAGGGGAAACAGTAAGGCTGATTGCTTGCTGAAAAAC 900

Qy        901 AGTAGGGCTGCTGGTTTGTTGCTAACGGAAAAAGGGGAGGTGTGTGTATGCTTTTGTGAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTAGGGCTGCTGGTTTGTTGCTAACGGAAAAAGGGGAGGTGTGTGTATGCTTTTGTGAC 960

Qy        961 TTTCCAGCGGAGGGGGTACAGTGCACGGCCAAGTTACAAG 1000
              ||||||||||||||||||||||||||||||||||||||||
Db        961 TTTCCAGCGGAGGGGGTACAGTGCACGGCCAAGTTACAAG 1000


RESULT 2
AAI64291
ID   AAI64291 standard; DNA; 128139 BP.
XX
AC   AAI64291;
XX
DT   22-APR-2002  (first entry)
XX
DE   RRV genome nucleotide sequence.
XX
KW   RRV; rhesus rhadinovirus; japanese macaque virus; multiple sclerosis;
KW   JMHV; cytostatic; antiasthmatic; antiallergic; dermatological;
KW   vulnerary; gene therapy; leucopenia; thrombocytopaenia;
KW   inflammatory disease; asthma; allergy; dermatitis; virus; ds.
XX
OS   Macaca mulatta rhadinovirus 17577.
XX
FH   Key             Location/Qualifiers
FT   CDS             1353..2674
FT                   /*tag= a
FT                   /product= "RRV R1"
FT   CDS             complement (2692..3258)
FT                   /*tag= b
FT                   /product= "dihydrofolate reductase"
FT                   /label= RRV_ORF2
FT                   /note= "has similarity to Kaposi's sarcoma-associated
FT                        virus (KSHV) open reading frame (ORF) 2"
FT   CDS             3676..5613
FT                   /*tag= c
FT                   /product= "complement binding protein"
FT                   /label= RRV_ORF4
FT                   /note= "has similarity to KSHV ORF4"
FT   CDS             6045..9443
FT                   /*tag= d
FT                   /product= "ssDNA binding protein"
FT                   /label= RRV_ORF6
FT                   /note= "has similarity to KSHV ORF6"
FT   CDS             9468..11528
FT                   /*tag= e
FT                   /product= "transport protein"
FT                   /label= RRV_ORF7
FT                   /note= "has similarity to KSHV ORF7"
FT   CDS             11515..14004
FT                   /*tag= f
FT                   /product= "glycoprotein B"
FT                   /label= RRV_ORF8
FT                   /note= "has similarity to KSHV ORF8"
FT   CDS             14122..17166
FT                   /*tag= g
FT                   /product= "DNA polymerase protein"
FT                   /label= RRV_ORF9
FT                   /note= "has similarity to KSHV ORF9"
FT   CDS             17261..18511
FT                   /*tag= h
FT                   /label= RRV_ORF10
FT                   /note= "has similarity to KSHV ORF10"
FT   CDS             18520..19749
FT                   /*tag= i
FT                   /label= RRV_ORF11
FT                   /note= "has similarity to KSHV ORF11"
FT   CDS             complement (19921..20544)
FT                   /*tag= j
FT                   /product= "RRV R2"
FT                   /label= RRV_ORF
FT                   /note= "has similarity to KSHV interleukin (IL)-6 gene"
FT   CDS             complement (20777..21778)
FT                   /*tag= k
FT                   /product= "thymidylate synthase"
FT                   /label= RRV_ORF70
FT                   /note= "has similarity to KSHV ORF70"
FT   CDS             complement (22245..22592)
FT                   /*tag= l
FT                   /product= "RRV R3"
FT                   /note= "has similarity to KSHV K4 viral MIP gene"
FT   CDS             26846..27409
```

```
FT                   /*tag= m
FT                   /product= "Bcl2-homologue"
FT                   /label= RRV_ORF16
FT                   /note= "has similarity to KSHV ORF16"
FT   CDS             complememt (27515..29125)
FT                   /*tag= n
FT                   /label= RRV_ORF17
FT                   /note= "has similarity to KSHV ORF17"
FT   CDS             28998..29897
FT                   /*tag= o
FT                   /label= RRV_ORF18
FT                   /note= "has similarity to KSHV ORF18"
FT   CDS             complement (29905..31548)
FT                   /*tag= p
FT                   /product= "tegument protein
FT                   /label= RRV_ORF19
FT                   /note= "has similarity to KSHV ORF19"
FT   CDS             complement (31043..32095)
FT                   /*tag= q
FT                   /label= RRV_ORF20
FT                   /note= "has similarity to KSHV ORF20"
FT   CDS             32094..33767
FT                   /*tag= r
FT                   /product= "thymidine kinase"
FT                   /label= RRV_ORF21
FT                   /note= "has similarity to KSHV ORF21"
FT   CDS             33754..35868
FT                   /*tag= s
FT                   /product= "glycoprotein H"
FT                   /label= RRV_ORF22
FT                   /note= "has similarity to KSHV ORF22"
FT   CDS             complement (35865..37073)
FT                   /*tag= t
FT                   /label= RRV_ORF23
FT                   /note= "has similarity to KSHV ORF23"
FT   CDS             complement (37123..39321)
FT                   /*tag= u
FT                   /label= RRV_ORF24
FT                   /note= "has similarity to KSHV ORF24"
FT   CDS             39323..43459
FT                   /*tag= v
FT                   /product= "major capsid protein"
FT                   /label= RRV_ORF25
FT                   /note= "has similarity to KSHV ORF25"
FT   CDS             43491..44408
FT                   /*tag= w
FT                   /product= "capsid protein"
FT                   /label= RRV_ORF26
FT                   /note= "has similarity to KSHV ORF26"
FT   CDS             44433..45242
FT                   /*tag= x
FT                   /label= RRV_ORF27
FT                   /note= "has similarity to KSHV ORF27"
FT   CDS             45408..45683
FT                   /*tag= y
FT                   /label= RRV_ORF28
FT                   /note= "has similarity to KSHV ORF28"
FT   CDS             complement (45733..46779)
FT                   /*tag= z
FT                   /label= RRV_ORF29b
FT                   /note= "has similarity to KSHV ORF29b"
FT   CDS             46905..47135
FT                   /*tag= aa
FT                   /label= RRV_ORF30
FT                   /note= "has similarity to KSHV ORF30"
FT   CDS             47093..47746
FT                   /*tag= ab
FT                   /label= RRV_ORF31
FT                   /note= "has similarity to KSHV ORF31"
FT   CDS             47683..49077
FT                   /*tag= ac
FT                   /label= RRV_ORF32
FT                   /note= "has similarity to KSHV ORF32"
FT   CDS             49049..50059
FT                   /*tag= ad
FT                   /label= RRV_ORF33
FT                   /note= "has similarity to KSHV ORF33"
FT   CDS             complement (49977..50960)
FT                   /*tag= ae
FT                   /label= RRV_ORF29a
FT                   /note= "has similarity to KSHV ORF29a"
FT   CDS             50959..51942
FT                   /*tag= af
FT                   /label= RRV_ORF34
FT                   /note= "has similarity to KSHV ORF34"
FT   CDS             51923..52372
FT                   /*tag= ag
FT                   /label= RRV_ORF35
FT                   /note= "has similarity to KSHV ORF35"
FT   CDS             52278..53585
FT                   /*tag= ah
FT                   /product= "kinase"
FT                   /label= RRV_ORF36
FT                   /note= "has similarity to KSHV ORF36"
FT   CDS             53566..55008
FT                   /*tag= ai
FT                   /product= "alkaline exonuclease"
FT                   /label= RRV_ORF37
FT                   /note= "has similarity to KSHV ORF37"
FT   CDS             54963..55172
FT                   /*tag= aj
FT                   /label= RRV_ORF38
FT                   /note= "has similarity to KSHV ORF38"
FT   CDS             complement (55255..56391)
FT                   /*tag= ak
FT                   /product= "glycoprotein M"
FT                   /label= RRV_ORF39
FT                   /note= "has similarity to KSHV ORF39"
FT   CDS             56526..57932
FT                   /*tag= al
FT                   /product= "helicase/primase"
FT                   /label= RRV_ORF40
FT                   /note= "has similarity to KSHV ORF40"
FT   CDS             57917..58528
FT                   /*tag= am
FT                   /product= "helicase/primase"
FT                   /label= RRV_ORF41
FT                   /note= "has similarity to KSHV ORF41"
FT   CDS             complement (58525..59343)
FT                   /*tag= an
FT                   /label= RRV_ORF42
FT                   /note= "has similarity to KSHV ORF42"
FT   CDS             complement (59297..61027)
FT                   /*tag= ao
FT                   /product= "capsid protein"
FT                   /label= RRV_ORF43
FT                   /note= "has similarity to KSHV ORF43"

  Query Match             79.2%;  Score 791.6;  DB 24;  Length 128139;
  Best Local Similarity   96.1%;  Pred. No. 5.6e-192;
  Matches  830;  Conservative    0;  Mismatches   14;  Indels   20;  Gaps    1;

Qy          1 GATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127289 GATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACA 127348

Qy         61 CCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127349 CCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 127408

Qy        121 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127409 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCT 127468

Qy        181 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     127469 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGC- 127527

Qy        241 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 300
                                 ||||||||||||||||||||||||||||||||||||||||
Db     127528 ------------------TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 127568

Qy        301 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127569 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 127628

Qy        361 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127629 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 127688

Qy        421 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127689 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 127748

Qy        481 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127749 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 127808

Qy        541 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127809 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 127868

Qy        601 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127869 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 127928

Qy        661 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127929 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGG 127988

Qy        721 GGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGGGCTGCCGGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127989 GGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGGGCTGCCGGC 128048

Qy        781 CAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACCCGGGCCTCTCCCCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     128049 CAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACCCGGGCCTCTCCCCC 128108

Qy        841 AGGCACAAAGCAGTAGGGCTGGCC 864
                |  |  || |     | ||||
Db     128109 GCCCGGACCGCCGGGCACCCGGCC 128132
```

```
RESULT 3
ABA41262
ID   ABA41262 standard; DNA; 576 BP.
XX
AC   ABA41262;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Probe #19728 for gene expression analysis in human heart cell sample.
XX
KW   Human; gene expression; heart; microarray; vascular system; probe;
KW   cardiovascular disease; hypertension; cardiac arrhythmia;
KW   congenital heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157274-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00666.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488899/53.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   hearts -
XX
PS   Claim 4; SEQ ID No 19728; 530pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes for
CC   measuring human gene expression in a sample derived from human heart. The
CC   present sequence is one such probe. The probes may be used for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from the human heart via microarrays. By measuring gene expression, the
CC   probes are useful for predicting, diagnosing, grading, staging,
CC   monitoring and prognosing diseases of the human heart and vascular system
CC   e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC   congenital heart disease.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 576 BP; 7 A; 60 C; 219 G; 290 T; 0 other;
```

Query Match 19.3%; Score 192.8; DB 22; Length 576;
Best Local Similarity 59.2%; Pred. No. 1.1e-39;
Matches 329; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

```
Qy        180 TGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGC 239
              || ||| |   | |  |||||| |  |  |   |   |  || | |  || | |   ||
Db          8 TGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTTGGCTGGTTGCTTGGT 67

Qy        240 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 299
              || | ||| |  ||||||| ||| | | | |  ||||||||||||| ||| || ||||||||
Db         68 TGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 127

Qy        300 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 359
              |||| ||| || ||  || ||| | | | |  |||||| ||| | | | || ||||||||
Db        128 TGCTTGCTTGCTTGGCTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGCTTGCTTGCT 187

Qy        360 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              || | ||| || |  ||| |||||  | | || |||||  ||| | ||| |  ||||||||
Db        188 TGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCT 247

Qy        420 GAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGAT 479
                |  |   |   || || |||||||  |  |  |   |   |  ||  || | | ||| |
Db        248 TCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGCT 307

Qy        480 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 539
                |   || |||| || |||| |   || || | || || | ||  || |||| || || |
Db        308 TGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGTTTGGTTGCTTGGTTGGT 367

Qy        540 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 599
                |   || || | |  || | |   || |  | |  |||| |    | |||| |  || |
Db        368 TGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTTGTTGCTTGCTTGGT 427

Qy        600 AGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 659
                  || || | || || | |   || || | || || | |   || || |  | || |
Db        428 TCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTCGTTGGTTGAT 487

Qy        660 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTG 719
                |   || || | || |||| |   || || ||| |   ||| ||| ||| ||||| |
Db        488 TGCTTGGTTGGTTGGTTGCTTGGTTGGTTGGCTGGTTCATTGGTTGGTTGGTTGCTTGGT 547

Qy        720 GGGCCGCTTGCCTGCT 735
                |   | |||  || |
Db        548 TGTTTGGTTGGTTGGT 563
```

```
RESULT 4
AAI57480
ID   AAI57480 standard; DNA; 576 BP.
XX
AC   AAI57480;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #26166 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 26166; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 576 BP; 7 A; 60 C; 219 G; 290 T; 0 other;

  Query Match             19.3%;  Score 192.8;  DB 22;  Length 576;
  Best Local Similarity   59.2%;  Pred. No. 1.1e-39;
  Matches  329;  Conservative    0;  Mismatches  227;  Indels    0;  Gaps    0;

Qy        180 TGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGC 239
              || ||| |   | |  |||||| |  |  |    |  || | |  || | |   ||
Db          8 TGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTTGGCTGGTTGCTTGGT 67

Qy        240 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 299
              || | ||| |  |||||| ||| | | | |  ||||||||||| ||| || ||||||||
Db         68 TGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 127

Qy        300 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 359
              |||| ||| || ||  || ||| | | | |  |||||| ||| | | | || ||||||||
Db        128 TGCTTGCTTGCTTGGCTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGCTTGCTTGCT 187

Qy        360 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              || | ||| || |  ||| ||||| | | || |||||  ||| | ||| |  ||||||||
Db        188 TGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCT 247

Qy        420 GAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGAT 479
                |  |   |  || || ||||||| |  |  |   |  |  ||  || | | ||| |
Db        248 TCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCT 307

Qy        480 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 539
              |   || |||| || ||||  |   || || | || || | ||  || |||| || || |
Db        308 TGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGTTTGGTTGCTTGGTTGGT 367

Qy        540 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 599
              |   || || | |  || | |   || |  | |  |||| |    | |||| |  || |
Db        368 TGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTTGTTGCTTGCTTGGT 427

Qy        600 AGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 659
                || || | || || | |   || || | || || | |   || || |  | || |
Db        428 TCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTCGTTGAT 487

Qy        660 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTG 719
              |   || || | || |||| |   || || ||| |   ||| ||| ||| ||||| |
Db        488 TGCTTGGTTGGTTGGTTGCTTGGTTGGTTGGCTGGTTCATTGGTTGGTTGGTTGCTTGGT 547

Qy        720 GGGCCGCTTGCCTGCT 735
              |   | |||  || |
Db        548 TGTTTGGTTGGTTGGT 563


RESULT 5
ABA31139
ID   ABA31139 standard; DNA; 555 BP.
XX
AC   ABA31139;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Probe #9605 for gene expression analysis in human heart cell sample.
XX
KW   Human; gene expression; heart; microarray; vascular system; probe;
KW   cardiovascular disease; hypertension; cardiac arrhythmia;
KW   congenital heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157274-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00666.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488899/53.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   hearts -
XX
PS   Claim 1; SEQ ID No 9605; 530pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes for
CC   measuring human gene expression in a sample derived from human heart. The
CC   present sequence is one such probe. The probes may be used for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from the human heart via microarrays. By measuring gene expression, the
CC   probes are useful for predicting, diagnosing, grading, staging,
CC   monitoring and prognosing diseases of the human heart and vascular system
CC   e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC   congenital heart disease.
CC   Note: The sequence data for this patent did not form part of the printed
```

```
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 555 BP; 18 A; 59 C; 201 G; 277 T; 0 other;

  Query Match             18.1%;  Score 180.6;  DB 22;  Length 555;
  Best Local Similarity   60.1%;  Pred. No. 1.4e-36;
  Matches  300;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy        217 GGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 276
              || |||| | ||| | |    | || | ||| |  || |||||||| | ||| |  ||  |
Db         56 GGTTGCTTGGTTGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTTGGCT 115

Qy        277 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 336
              | ||||| | | |  |||||| ||||| | | |  || ||| ||||| ||| || |||||
Db        116 GGTTGCTTGGTTGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTGCTT 175

Qy        337 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 396
              ||||||| ||| || |||||||||||   | | |  || ||| ||||| | | |  || ||
Db        176 GCTTGCTTGCTTGCTTGCTTGCTTGGCTGGTTGGTTGGTTGCTTGGTTGGTTGGTT 235

Qy        397 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACG 456
              ||||||| ||| |  ||||||||     |   |   || |||||||||  |  |  |
Db        236 GCTTGCTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCATTGGTTGCTT 295

Qy        457 GTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 516
              |   || ||| |  | ||||| | |   || |||| || || | |   || || | || |
Db        296 GGTTGCTTGCTTCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTTGGTT 355

Qy        517 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCT 576
              | | |   |  || | || |||| |   |  || | || || | |   || | | || |
Db        356 GGTTGGTTGCTTGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGTTTGGTT 415

Qy        577 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCT 636
              ||| |   || || | || || | |   || || | || || | |   || ||  || |
Db        416 GCTTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTTGTT 475

Qy        637 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTG 696
              ||| |   || |  | || |  | |   || || | || || | |   || ||  || |
Db        476 GCTTGCTTGGTTCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 535

Qy        697 GCTTGCTTGCTTGCTTGCT 715
              | ||  ||| |||||| |
Db        536 GGTTCGTTGATTGCTTGGT 554


RESULT 6
AAI44102
ID   AAI44102 standard; DNA; 555 BP.
XX
AC   AAI44102;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #12788 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 12788; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 555 BP; 18 A; 59 C; 201 G; 277 T; 0 other;

  Query Match             18.1%;  Score 180.6;  DB 22;  Length 555;
  Best Local Similarity   60.1%;  Pred. No. 1.4e-36;
  Matches  300;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy        217 GGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 276
              || |||| | ||| | |    | || | ||| |  || |||||||| | ||| |  ||  |
Db         56 GGTTGCTTGGTTGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTTGGCT 115

Qy        277 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 336
              | ||||| | | |  |||||| ||||| | | |  || ||| ||||| ||| || |||||
Db        116 GGTTGCTTGGTTGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTGCTT 175

Qy        337 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 396
              ||||||| ||| || |||||||||||   | | |  || ||| ||||| | | |  || ||
Db        176 GCTTGCTTGCTTGCTTGCTTGCTTGGCTGGTTGGTTGGTTGCTTGGTTGGTTGGTT 235

Qy        397 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACG 456
              ||||||| ||| |  ||||||||     |   |   || |||||||||  |  |  |
Db        236 GCTTGCTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCATTGGTTGCTT 295

Qy        457 GTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 516
              |   || ||| |  | ||||| | |   || |||| || || | |   || || | || |
Db        296 GGTTGCTTGCTTCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTTGGTT 355

Qy        517 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCT 576
              | | |   |  || | || |||| |   |  || | || || | |   || | | || |
Db        356 GGTTGGTTGCTTGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGTTTGGTT 415

Qy        577 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCT 636
              ||| |   || || | || || | |   || || | || || | |   || ||  || |
Db        416 GCTTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTTGTT 475

Qy        637 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTG 696
              ||| |   || |  | || |  | |   || || | || || | |   || ||  || |
Db        476 GCTTGCTTGGTTCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 535

Qy        697 GCTTGCTTGCTTGCTTGCT 715
              | ||  ||| |||||| |
Db        536 GGTTCGTTGATTGCTTGGT 554


RESULT 7
AAC64754/c
ID   AAC64754 standard; DNA; 133719 BP.
XX
AC   AAC64754;
XX
DT   28-FEB-2001  (first entry)
XX
DE   Macaca mulatta rhadinovirus 17577 (RRV) genome sequence SEQ ID NO:1.
XX
KW   Macaca mulatta rhadinovirus 17577; RRV; rhesus macaque rhadinovirus;
```

```
KW   genome; Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6;
KW   IL-6; macrophage inflammatory protein; MIP; diagnosis; vaccine;
KW   cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma;
KW   lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy;
KW   splenomegaly; hypergammaglobinulinaemia; autoimmune haemolytic anaemia;
KW   ds.
XX
OS   Macaca mulatta rhadinovirus 17577.
XX
PN   WO200028040-A2.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26260.
XX
PR   06-NOV-1998;   98US-0107507.
PR   20-NOV-1998;   98US-0109409.
XX
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Wong SW,  Axthelm MK,  Searles RP;
XX
DR   WPI; 2000-376552/32.
XX
PT   New rhesus rhadino virus for producing non-human primate model useful
PT   for testing potential treatments and efficacy of the candidate vaccine
PT   for conditions associated with RRV infection  -
XX
PS   Claim 2; Page 83-122; 141pp; English.
XX
CC   The present invention describes a novel rhesus macaque rhadinovirus
CC   called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the
CC   RRV genome sequence, and AAB53123 to AAB53204 represent the proteins
CC   encoded by the genome sequence. The present invention also specifically
CC   claims the individual open reading frame (ORF) nucleotide sequences from
CC   the genome which encode the individual proteins, but these sequences are
CC   not given. A non-human animal infected with RRV can be used for testing
CC   the efficacy of drug in the treatment of condition associated with
CC   infection with RRV such as Kaposi's sarcoma, lymphoproliferative
CC   disorders, B-cell hyperplasia, lymphadenopathy, splenomegaly,
CC   hypergammaglobinulinaemia or autoimmune haemolytic anaemia, by
CC   administering the drug to a immuno-compromised non-human primate
CC   preferably Rhesus macaque monkey obtained by as a result of infection
CC   by Simian Immunodeficiency Virus (SIV). RRV is useful for producing
CC   non-human primate model for testing potential treatments for conditions
CC   associated with RRV infection. It is also useful for testing the
CC   efficacy of the candidate vaccine against RRV infection or conditions
CC   associated with its infection by administering the vaccine to the
CC   subject capable of infection with RRV, inoculating the subject with RRV
CC   and observing the effect of vaccine. AAC64755 to AAC64765 and AAB53205
CC   to AAB53213 represent sequence used in the exemplification of the
CC   present invention.
XX
SQ   Sequence 133719 BP; 32746 A; 35648 C; 34521 G; 30804 T; 0 other;

  Query Match             14.5%;  Score 144.8;  DB 21;  Length 133719;
  Best Local Similarity   51.9%;  Pred. No. 9.8e-27;
  Matches  326;  Conservative    0;  Mismatches  302;  Indels    0;  Gaps    0;

Qy        108 AGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGG 167
              || ||||  ||   | | || ||  ||  ||      | ||| ||||   | | ||
Db        735 AGCAGGCAAGCGGCCCCACTAGCAAGCAAGCAAGCAAGCCAGCAGGCAGCCCTACTAGCA 676

Qy        168 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCT 227
              |  ||| || ||| || ||   |  ||   | | ||    |  ||| ||  || || ||
Db        675 GCCCAGCAGCCCTACTAGCAGCCCAGCAGCCCTACTAGCAGCCCAGCAGCCCTACTAGCA 616

Qy        228 TGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 287
                | || ||  || || || | || ||   | | || || | || |    | | || ||
Db        615 GCCCAGGAGCCCTACTAGCAGCCCAGCAGCCCTACTAGCAGCCCAGGAGCCCTACTAGCA 556

Qy        288 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 347
              | || ||   | | || || | || ||   | | || || | || ||   | | || ||

Db        555 GCCCAGCAGCCCTACTAGCAGCCCAGCAGCCCTACTAGCAGCCCAGCAGCCCTACTAGCA 496

Qy        348 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 407
              | || ||   | | ||  |  |||  |  ||  ||   | | ||   | ||  ||  ||
Db        495 GCCCAGCAGCCCTACTATCAAGCCATCAGGCAGGCGTACCGTCCCCTTAGCAAGCAAGCA 436

Qy        408 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCC 467
              | ||| ||  |       |  || ||| | ||  ||  || |   || | |   ||  ||
Db        435 GCCCTACTGTCCCCTCAGCAAGCAGGCCAGCGAGCAAGCAAGCAGGCCAGCGAGCAAGCA 376

Qy        468 TGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 527
                |  |||  |  || | |   || || | || || | |   || || | || || | |
Db        375 AGCAGGCCAGCGAGCAAGCAAGCAGGCCAGCGAGCAAGCAAGCAGGCCAGCGAGCAAGCA 316

Qy        528 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGC 587
              || || | || || | |   || || | || || | |    | || | || || | |
Db        315 AGCAGGCCAGCGAGCAAGCAAGCAGGCCAGCGAGCAAGCAAGCAGGCCAGCGAGCAAGCA 256

Qy        588 TGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 647
              || || | || || ||  || || |   || || ||  || |||  |    || | |
Db        255 AGCAGGCCAGCGAGCAGCCCTACTAGCAAGCCAGCAGCCCTACTGTCCCCTCAGCAAGCA 196

Qy        648 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCT 707
              ||  |   || || ||| ||  |||  |    || | |   ||||  ||| |      ||
Db        195 GGCCAGCGAGCAAGCAGGCCTACTGTCCCCTCAGCAAGCAGGCCTACTGTCCCCTCAGCA 136

Qy        708 TGCTTGCTAGTGGGGCCGCTTGCCTGCT 735
              ||  || || | |   ||  |||| ||
Db        135 AGCAGGCCAGCGAGCAAGCAGGCCTACT 108


RESULT 8
AAI64291/c
ID   AAI64291 standard; DNA; 128139 BP.
XX
AC   AAI64291;
XX
DT   22-APR-2002  (first entry)
XX
DE   RRV genome nucleotide sequence.
XX
KW   RRV; rhesus rhadinovirus; japanese macaque virus; multiple sclerosis;
KW   JMHV; cytostatic; antiasthmatic; antiallergic; dermatological;
KW   vulnerary; gene therapy; leucopenia; thrombocytopaenia;
KW   inflammatory disease; asthma; allergy; dermatitis; virus; ds.
XX
OS   Macaca mulatta rhadinovirus 17577.
XX
FH   Key             Location/Qualifiers
FT   CDS             1353..2674
FT                   /*tag=  a
FT                   /product= "RRV R1"
FT   CDS             complement (2692..3258)
FT                   /*tag=  b
FT                   /product= "dihydrofolate reductase"
FT                   /label= RRV_ORF2
FT                   /note= "has similarity to Kaposi's sarcoma-associated
FT                         virus (KSHV) open reading frame (ORF) 2"
FT   CDS             3676..5613
FT                   /*tag=  c
FT                   /product= "complement binding protein"
FT                   /label= RRV_ORF4
FT                   /note= "has similarity to KSHV ORF4"
FT   CDS             6045..9443
FT                   /*tag=  d
FT                   /product= "ssDNA binding protein"
FT                   /label= RRV_ORF6
FT                   /note= "has similarity to KSHV ORF6"
FT   CDS             9468..11528
FT                   /*tag=  e
FT                   /product= "transport protein"
FT                   /label= RRV_ORF7
FT                   /note= "has similarity to KSHV ORF7"
```

```
FT   CDS             11515..14004
FT                   /*tag= f
FT                   /product= "glycoprotein B"
FT                   /label= RRV_ORF8
FT                   /note= "has similarity to KSHV ORF8"
FT   CDS             14122..17166
FT                   /*tag= g
FT                   /product= "DNA polymerase protein"
FT                   /label= RRV_ORF9
FT                   /note= "has similarity to KSHV ORF9"
FT   CDS             17261..18511
FT                   /*tag= h
FT                   /label= RRV_ORF10
FT                   /note= "has similarity to KSHV ORF10"
FT   CDS             18520..19749
FT                   /*tag= i
FT                   /label= RRV_ORF11
FT                   /note= "has similarity to KSHV ORF11"
FT   CDS             complement (19921..20544)
FT                   /*tag= j
FT                   /product= "RRV R2"
FT                   /label= RRV_ORF
FT                   /note= "has similarity to KSHV interleukin (IL)-6 gene"
FT   CDS             complement (20777..21778)
FT                   /*tag= k
FT                   /product= "thymidylate synthase"
FT                   /label= RRV_ORF70
FT                   /note= "has similarity to KSHV ORF70"
FT   CDS             complement (22245..22592)
FT                   /*tag= l
FT                   /product= "RRV R3"
FT                   /note= "has similarity to KSHV K4 viral MIP gene"
FT   CDS             26846..27409
FT                   /*tag= m
FT                   /product= "Bcl2-homologue"
FT                   /label= RRV_ORF16
FT                   /note= "has similarity to KSHV ORF16"
FT   CDS             complememt (27515..29125)
FT                   /*tag= n
FT                   /label= RRV_ORF17
FT                   /note= "has similarity to KSHV ORF17"
FT   CDS             28998..29897
FT                   /*tag= o
FT                   /label= RRV_ORF18
FT                   /note= "has similarity to KSHV ORF18"
FT   CDS             complement (29905..31548)
FT                   /*tag= p
FT                   /product= "tegument protein
FT                   /label= RRV_ORF19
FT                   /note= "has similarity to KSHV ORF19"
FT   CDS             complement (31043..32095)
FT                   /*tag= q
FT                   /label= RRV_ORF20
FT                   /note= "has similarity to KSHV ORF20"
FT   CDS             32094..33767
FT                   /*tag= r
FT                   /product= "thymidine kinase"
FT                   /label= RRV_ORF21
FT                   /note= "has similarity to KSHV ORF21"
FT   CDS             33754..35868
FT                   /*tag= s
FT                   /product= "glycoprotein H"
FT                   /label= RRV_ORF22
FT                   /note= "has similarity to KSHV ORF22"
FT   CDS             complement (35865..37073)
FT                   /*tag= t
FT                   /label= RRV_ORF23
FT                   /note= "has similarity to KSHV ORF23"
FT   CDS             complement (37123..39321)
FT                   /*tag= u
FT                   /label= RRV_ORF24
FT                   /note= "has similarity to KSHV ORF24"
FT   CDS             39323..43459
FT                   /*tag= v
FT                   /product= "major capsid protein"
FT                   /label= RRV_ORF25
FT                   /note= "has similarity to KSHV ORF25"
FT   CDS             43491..44408
FT                   /*tag= w
FT                   /product= "capsid protein"
FT                   /label= RRV_ORF26
FT                   /note= "has similarity to KSHV ORF26"
FT   CDS             44433..45242
FT                   /*tag= x
FT                   /label= RRV_ORF27
FT                   /note= "has similarity to KSHV ORF27"
FT   CDS             45408..45683
FT                   /*tag= y
FT                   /label= RRV_ORF28
FT                   /note= "has similarity to KSHV ORF28"
FT   CDS             complement (45733..46779)
FT                   /*tag= z
FT                   /label= RRV_ORF29b
FT                   /note= "has similarity to KSHV ORF29b"
FT   CDS             46905..47135
FT                   /*tag= aa
FT                   /label= RRV_ORF30
FT                   /note= "has similarity to KSHV ORF30"
FT   CDS             47093..47746
FT                   /*tag= ab
FT                   /label= RRV_ORF31
FT                   /note= "has similarity to KSHV ORF31"
FT   CDS             47683..49077
FT                   /*tag= ac
FT                   /label= RRV_ORF32
FT                   /note= "has similarity to KSHV ORF32"
FT   CDS             49049..50059
FT                   /*tag= ad
FT                   /label= RRV_ORF33
FT                   /note= "has similarity to KSHV ORF33"
FT   CDS             complement (49977..50960)
FT                   /*tag= ae
FT                   /label= RRV_ORF29a
FT                   /note= "has similarity to KSHV ORF29a"
FT   CDS             50959..51942
FT                   /*tag= af
FT                   /label= RRV_ORF34
FT                   /note= "has similarity to KSHV ORF34"
FT   CDS             51923..52372
FT                   /*tag= ag
FT                   /label= RRV_ORF35
FT                   /note= "has similarity to KSHV ORF35"
FT   CDS             52278..53585
FT                   /*tag= ah
FT                   /product= "kinase"
FT                   /label= RRV_ORF36
FT                   /note= "has similarity to KSHV ORF36"
FT   CDS             53566..55008
FT                   /*tag= ai
FT                   /product= "alkaline exonuclease"
FT                   /label= RRV_ORF37
FT                   /note= "has similarity to KSHV ORF37"
FT   CDS             54963..55172
FT                   /*tag= aj
FT                   /label= RRV_ORF38
FT                   /note= "has similarity to KSHV ORF38"
FT   CDS             complement (55255..56391)
FT                   /*tag= ak
FT                   /product= "glycoprotein M"
FT                   /label= RRV_ORF39
FT                   /note= "has similarity to KSHV ORF39"
FT   CDS             56526..57932
FT                   /*tag= al
FT                   /product= "helicase/primase"
FT                   /label= RRV_ORF40
FT                   /note= "has similarity to KSHV ORF40"
```

```
FT   CDS.          •    57917..58528
FT                      /*tag=  am
FT                      /product= "helicase/primase"
FT                      /label= RRV_ORF41
FT                      /note= "has similarity to KSHV ORF41"
FT   CDS                complement (58525..59343)
FT                      /*tag=  an
FT                      /label= RRV_ORF42
FT                      /note= "has similarity to KSHV ORF42"
FT   CDS                complement (59297..61027)
FT                      /*tag=  ao
FT                      /product= "capsid protein"
FT                      /label= RRV_ORF43
FT                      /note= "has similarity to KSHV ORF43"

  Query Match             13.4%;  Score 134.2;  DB 24;  Length 128139;
  Best Local Similarity   51.6%;  Pred. No. 4.9e-24;
  Matches  333;  Conservative    0;  Mismatches  308;  Indels    4;  Gaps    1;

Qy        108 AGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGG 167
              || ||||  ||   | | || ||  ||  ||       | ||| ||||   | | ||
Db     128003 AGCAGGCAAGCGGCCCCACTAGCAAGCAAGCAAGCAAGCCAGCAGGCAGCCCTACTAGCA 127944

Qy        168 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCT 227
              |  ||| || ||| || ||   |  ||   | | ||    |  ||| ||  || || ||
Db     127943 GCCCAGCAGCCCTACTAGCAGCCCAGCAGCCCTACTAGCAGCCCAGCAGCCCTACTAGCA 127884

Qy        228 TGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 287
              | || ||  || || || | || ||   | | || || | || |    | | || ||
Db     127883 GCCCAGGAGCCCTACTAGCAGCCCAGCAGCCCTACTAGCAGCCCAGGAGCCCTACTAGCA 127824

Qy        288 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 347
              | || ||   | | || || | || ||   | | || || | || ||   | | || ||
Db     127823 GCCCAGCAGCCCTACTAGCAGCCCAGCAGCCCTACTAGCAGCCCAGCAGCCCTACTAGCA 127764

Qy        348 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 407
              | || ||   | | ||  |  |||  |  ||  ||   | | ||   | ||  ||  ||
Db     127763 GCCCAGCAGCCCTACTATCAAGCCATCAGGCAGGCGTACCGTCCCCTTAGCAAGCAAGCA 127704

Qy        408 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCC 467
              | ||| ||  |      |  || ||| | ||  ||  || |   || | |   || ||
Db     127703 GCCCTACTGTCCCCTCAGCAAGCAGGCCAGCGAGCAAGCAAGCAGGCCAGCGAGCAAGCA 127644

Qy        468 TGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 527
              |  |||  |  || | |   || || | || || | |   || || | || || | |
Db     127643 AGCAGGCCAGCGAGCAAGCAAGCAGGCCAGCGAGCAAGCAAGCAGGCCAGCGAGCAAGCA 127584

Qy        528 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGC 587
              || || | || || | |   || || | || || | |   | || | || || ||  |
Db     127583 AGCAGGCCAGCGAGCAAGCAAGCAGGCCAGCGAGCAAGCAAGCAGGCCAGCGAGCAGCCC 127524

Qy        588 TGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 647
              | || |   || || ||  || |||  |    || | |  ||  |   || || ||| |
Db     127523 TACTAGCAAGCCAGCAGCCCTACTGTCCCCTCAGCAAGCAGGCCAGCGAGCAAGCAGGCC 127464

Qy        648 TGCTG----GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCT 703
              | |||      | || || ||| || |||  |    || | |   ||| ||  ||  ||
Db     127463 TACTGTCCCCTCAGCAAGCAGGCCTACTGTCCCCTCAGCAAGCAGGCCAGCGAGCAAGCA 127404

Qy        704 TGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTG 748
              || | ||  |   |  |  ||  ||| || |  |||   || |
Db     127403 GGCCTACTGTCCCCTCAGCAAGCAGGCCAGCGAGCAGGCAGGCAG 127359


RESULT 9
AAV73802/c
ID   AAV73802 standard; DNA; 35100 BP.
XX
AC   AAV73802;
XX
DT   25-FEB-1999  (first entry)
XX
DE   KSHV LUR DNA (nucleotides 1-35,100).
XX
KW   Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHFR; Bcl-2;
KW   dihydrofolate reductase; LUR; long unique region; vaccine; prophylaxis;
KW   diagnosis; treatment; HHV8; complement binding protein; v-CBP; SSBP;
KW   ssDNA binding protein; transport protein; glycoprotein B; pol; vIL-6;
KW   DNA polymerase; viral interleukin-6; BHV4-IE1 I; thymidylate synthase;
KW   vMIP-II; BHV4-IE1 II; vMIP-I; capsid protein I; tegument protein I; ds.
XX
OS   Kaposi's sarcoma-associated herpesvirus.
XX
PN   US5849564-A.
XX
PD   15-DEC-1998.
XX
PF   29-NOV-1996;   96US-0770379.
XX
PR   29-NOV-1996;   96US-0770379.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Bohenzky RA,  Chang Y,  Edelman IS,  Moore PS,  Russo JJ;
XX
DR   WPI; 1999-069741/06.
XX
PT   Kaposi's sarcoma-associated herpes virus nucleic acid - encodes
PT   di:hydro:folate reductase and is useful for treatment, prophylaxis
PT   or diagnosis of Kaposi's sarcoma
XX
PS   Disclosure; Column 67-96; 109pp; English.
XX
CC   This sequence is a fragment of the Kaposi's sarcoma-associated
CC   herpesvirus (KSHV) LUR (long unique region). This fragment contains
CC   coding regions for K1, ORF4 which encodes the complement binding protein
CC   v-CBP, ORF6 which encodes a ssDNA binding protein (SSBP), ORF7 which
CC   encodes a transport protein, ORF8 which encodes glycoprotein B, ORF9
CC   which encodes DNA polymerase (pol), ORF10, OFR11, K2 which encodes viral
CC   interleukin-6 (vIL-6), ORF02 which encodes dihydrofolate reductase
CC   (DHFR), K3 which encodes BHV4-IE1 I, ORF70 which encodes Thymidylate
CC   synthase, K4 which encodes vMIP-II, K5 which encodes BHV4-IE1 II, K6
CC   which encodes vMIP-I, K7, ORF16 which encodes Bcl-2, ORF17 which encodes
CC   capsid protein I, ORF18 and ORF19 which encodes Tegument protein I.
CC   KSHV is a new human Herpesvirus (HHV8) believed to cause Kaposi's sarcoma
CC   (KS) which is the most common form of neoplasm occurring in persons with
CC   acquired immune deficiency syndrome (AIDS). The DHFR protein is useful
CC   for vaccination, prophylaxis, diagnosis and treatment of a subject with
CC   Kaposi's sarcoma and for detecting expression of a DNA virus associated
CC   with Kaposi's sarcoma in a cell.
XX
SQ   Sequence 35100 BP; 8703 A; 9395 C; 8921 G; 8081 T; 0 other;

  Query Match             12.8%;  Score 127.6;  DB 20;  Length 35100;
  Best Local Similarity   66.9%;  Pred. No. 1.6e-22;
  Matches  198;  Conservative    0;  Mismatches   94;  Indels    4;  Gaps    1;

Qy        465 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 524
              | ||  |||  || |    ||||| ||||| ||||    ||||| ||||| ||||    |
Db     24611 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24552

Qy        525 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 584
              |||| ||||| ||||    ||||| ||||| ||||    |||||||||| ||||    |
Db     24551 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24492

Qy        585 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 644
              |||| ||||| ||||    |||||||||| ||||    ||||| ||||| ||||    |
Db     24491 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24432

Qy        645 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT----GCCTGCTGGCTT 700
              |||| ||||| ||||    ||||| ||||| ||||    |||||    |  ||||| |
Db     24431 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24372

Qy        701 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTG 756
              |  | || |  ||||   |||||  || |  |||| |  ||||| ||| |  ||||
Db     24371 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTG 24316
```

```
RESULT 10
AAV19941/c
ID   AAV19941 standard; DNA; 137507 BP.
XX
AC   AAV19941;
XX
DT   03-AUG-1998  (first entry)
XX
DE   KSHV long unique coding region and terminal repeat.
XX
KW   KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II;
KW   interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;
KW   complement-binding protein; glycoprotein; capsid protein IV; infection;
KW   immediate early protein; Kaposi's sarcoma; protective vaccine; lymphoma;
KW   lymphoproliferative disease; leukaemia; splenomegaly; mycosis fungoides;
KW   HIV immune status; anti-inflammatory agent; therapy; ds.
XX
OS   Kaposi's sarcoma-associated herpes virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1142..2794
FT                   /*tag= a
FT                   /product= complement-binding protein
FT   CDS             8699..11236
FT                   /*tag= b
FT                   /product= glycoprotein B
FT   CDS             complement (17261..17875)
FT                   /*tag= c
FT                   /product= interleukin 6
FT   CDS             complement (21548..21832)
FT                   /*tag= d
FT                   /product= macrophage inflammatory protein II
FT   CDS             complement (27137..27424)
FT                   /*tag= e
FT                   /product= interferon regulatory factor 1
FT   CDS             28661..29741
FT                   /*tag= f
FT                   /product= protein T1.1
FT   CDS             complement (58976..60175)
FT                   /*tag= g
FT                   /product= glycoprotein M
FT   CDS             complement (69412..69915)
FT                   /*tag= h
FT                   /product= glycoprotein L
FT   CDS             complement (88410..88910)
FT                   /*tag= i
FT                   /product= interferon regulatory factor 2
FT   CDS             89600..90541
FT                   /*tag= j
FT                   /product= interferon regulatory factor 3
FT   CDS             90173..90643
FT                   /*tag= k
FT                   /product= glycoprotein X
FT   CDS             complement (93636..94127)
FT                   /*tag= l
FT                   /product= interferon regulatory factor 4
FT   CDS             complement (111931..112443)
FT                   /*tag= m
FT                   /product= capsid protein IV
FT   CDS             complement (123808..127296)
FT                   /*tag= n
FT                   /product= immediate early protein
XX
PN   WO9804576-A1.
XX
PD   05-FEB-1998.
XX
PF   22-JUL-1997;    97WO-US13346.
XX
PR   29-NOV-1996;    96US-0757669.
PR   25-JUL-1996;    96US-0686243.
PR   25-JUL-1996;    96US-0686349.
PR   25-JUL-1996;    96US-0686350.
PR   25-JUL-1996;    96US-0687253.
PR   25-JUL-1996;    96US-0688814.
PR   05-SEP-1996;    96US-0708678.
PR   10-OCT-1996;    96US-0728323.
PR   13-NOV-1996;    96US-0747887.
PR   13-NOV-1996;    96US-0748640.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Bohenzky RA,  Chang Y,  Edelman IS,  Moore PS,  Russo JJ;
XX
DR   WPI; 1998-130615/12.
XX
PT   New nucleic acid encoding Kaposi's sarcoma associated herpes virus
PT   proteins - useful for, e.g. detecting levels of HHV8 in, and
PT   preparation of vaccines for treatment of, HIV patients
XX
PS   Example 2; Page 135-203; 230pp; English.
XX
CC   This sequence represents the long unique region and terminal repeat of
CC   the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known
CC   as human herpes virus 8 (HHV8). This sequence contains the DNAs of the
CC   invention which encode KSHV polypeptides selected from: (a) viral
CC   macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);
CC   (c) viral IRF 1; (d) complement-binding protein; glycoproteins B, M or L;
CC   (d) capsid protein IV encoded by ORF65; and (e) immediate early protein
CC   encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded
CC   by it, and antibodies (Ab) specific for the proteins are useful for
CC   detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body
CC   fluids or tissue samples. HHV8 infections can be treated with antisense
CC   or triplex forming molecules or agents that bind specifically to the
CC   protein. Ab may be used for prophylaxis or treatment of HHV8 infection,
CC   while the protein can be used in protective vaccines. Ab may also be used
CC   to differentiate between lymphomas, and HHV8 may be implicated in many
CC   other lymphoproliferative diseases such as lymphomas, leukaemia,
CC   splenomegaly and mycosis fungoides. Cells and animals containing the
CC   nucleic acid are useful for drug screening. HHV8-derived peptides can be
CC   used as targets for antiviral drugs, e.g. dihydrofolate reductase gene
CC   can be inhibited with methotrexate. These can also be used to determine
CC   the immune status of a patient infected with HIV. HHV8 derived protein
CC   viral MIP III may be used as an anti-inflammatory agent for,
CC   e.g. treating rheumatoid arthritis. This sequence is stated as containing
CC   81 open reading frames.
XX
SQ   Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T; 0 other;

  Query Match           12.8%;  Score 127.6;  DB 19;  Length 137507;
  Best Local Similarity   66.9%;  Pred. No. 2.4e-22;
  Matches  198;  Conservative    0;  Mismatches   94;  Indels    4;  Gaps    1;

Qy        465 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 524
              | ||  |||  || |    ||||| ||||| ||||    ||||| ||||| ||||  |
Db      24611 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24552

Qy        525 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 584
              |||| ||||| ||||    ||||| ||||| ||||    |||||||||||| ||||    |
Db      24551 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24492

Qy        585 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 644
              |||| ||||| ||||    |||||||||||| ||||    ||||| ||||| ||||    |
Db      24491 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24432

Qy        645 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT----GCCTGCTGGCTT 700
              |||| ||||| ||||    ||||| ||||| ||||    |||||    |  ||||| |
Db      24431 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24372

Qy        701 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTG 756
              |  | || |  |||   |||||  || |  |||| |  ||||| ||| |  ||||
Db      24371 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTG 24316
```

```
RESULT 11
AAQ95095
ID   AAQ95095 standard; DNA; 385 BP.
XX
AC   AAQ95095;
XX
DT   02-OCT-1995  (first entry)
XX
DE   Configuration detecting nucleic acid method target sequence.
XX
KW   Probe; PCR; terminus; amplify; primer; configuration; ss.
XX
OS   Synthetic.
XX
PN   JP06343498-A.
XX
PD   20-DEC-1994.
XX
PF   03-JUN-1993;   93JP-0133640.
XX
PR   03-JUN-1993;   93JP-0133640.
XX
PA   (CANO ) CANON KK.
XX
DR   WPI; 1995-069322/10.
XX
PT   Nucleic acid probe - and method for detecting nucleic acid
XX
PS   Example 1; Column 5; 6pp; Japanese.
XX
CC   The sequence of the PCR amplified target region detected by probes
CC   (AAQ95091-4).  The probes used in a method for the detection of a
CC   nucleic acid target sequence which has PCR-controlled termini and can be
CC   amplified by PCR.  The probes can be used to detect configuration.
XX
SQ   Sequence 385 BP; 93 A; 97 C; 98 G; 97 T; 0 other;

  Query Match             10.3%;  Score 103.4;  DB 16;  Length 385;
  Best Local Similarity   57.1%;  Pred. No. 6.5e-17;
  Matches  188;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy        116 TGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTA 175
              ||| |||  ||  || ||| |||      |   | | || ||| |||    | |  ||||
Db         43 TGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCGCTAGAGTAAGTA 102

Qy        176 GGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTA 235
              |   ||| |||  ||  || ||| |||      |   | | |  |||  || |||  | |
Db        103 GTTATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCA 162

Qy        236 GGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 295
                | |||  ||  || ||| ||| |||  ||  || ||||   |   |     || |||
Db        163 TGCATGCATGCATGCATGCATGCATGCATGCATGCTATATATATATAATGCATGCA 222

Qy        296 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 355
              ||| |||  ||  || ||| ||| |||  ||  || ||| ||| |||  ||  || |||
Db        223 TGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCA 282

Qy        356 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 415
              ||| |||  ||  || ||| ||| |||  ||  || ||| |||  ||  || |||
Db        283 TGCATGCATGCATGCATGCATGCGCGCGCGCGCGCATGCATGCATGCATGCA 342

Qy        416 TGCTGAGGGGACAGTAGGGCTGCTTGCTT 444
              |||      |   | | |  ||| ||| |
Db        343 TGCATGCATGCATGCATGCATGCGT 371


RESULT 12
AAQ84040
ID   AAQ84040 standard; DNA; 344 BP.
XX
AC   AAQ84040;
XX
DT   03-OCT-1995  (first entry)
XX
DE   Plasmid pCU19.
XX
KW   Plasmid pUC19; DNA detection metod; ss.
XX
OS   Synthetic.
XX
PN   JP06343499-A.
XX
PD   20-DEC-1994.
XX
PF   04-JUN-1993;   93JP-0134615.
XX
PR   04-JUN-1993;   93JP-0134615.
XX
PA   (CANO ) CANON KK.
XX
DR   WPI; 1995-069323/10.
XX
PT   Method for detecting nucleic acid - using nucleic acid probe
PT   and PCR amplification
XX
PS   Example 1; Page 4; 9pp; Japanese.
XX
CC   AAQ84035 is the plasmid pUC19, used to demonstrate a new method for
CC   the detection of DNA.
XX
SQ   Sequence 344 BP; 84 A; 85 C; 86 G; 89 T; 0 other;

  Query Match             10.2%;  Score 101.8;  DB 16;  Length 344;
  Best Local Similarity   56.8%;  Pred. No. 1.6e-16;
  Matches  187;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy        116 TGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTA 175
              ||| |||  ||  || ||| |||      |   | | || ||| |||    | |  ||||
Db          2 TGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCGCTAGAGTAAGTA 61

Qy        176 GGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTA 235
              |   ||| |||  ||  || ||| |||      |   | | |  |||  || |||  | |
Db         62 GTTATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCA 121

Qy        236 GGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 295
                | |||   |  || ||| ||| |||  ||  || ||||   |   |     || |||
Db        122 TGCATGCATCCATGCATGCATGCATGCATGCATGCTATATATATATAATGCATGCA 181

Qy        296 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 355
              ||| |||  ||  || ||| ||| |||  ||  || ||| ||| |||  ||  || |||
Db        182 TGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCA 241

Qy        356 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 415
              ||| |||  ||  || ||| ||| |||  ||  || ||| |||  ||  || |||
Db        242 TGCATGCATGCATGCATGCATGCGCGCGCGCGCGCATGCATGCATGCATGCA 301

Qy        416 TGCTGAGGGGACAGTAGGGCTGCTTGCTT 444
              |||      |   | | |  ||| ||| |
Db        302 TGCATGCATGCATGCATGCATGCGT 330


RESULT 13
AAK79692
ID   AAK79692 standard; DNA; 14212 BP.
XX
AC   AAK79692;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34504.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
```

PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;

```
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 34504; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 14212 BP; 3809 A; 3421 C; 3440 G; 3542 T; 0 other;

  Query Match             10.0%;  Score 99.6;  DB 22;  Length 14212;
  Best Local Similarity   51.1%;  Pred. No. 1.8e-15;
  Matches  286;  Conservative    0;  Mismatches  269;  Indels    5;  Gaps    2;

Qy        137 GCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 196
              | |||  ||  ||||||   | | | ||   ||   |||||  || ||  | | | | ||
Db       6620 GTTGACTGGTTAGTAGGTTGGTTAGTTGGTTGGCTGGTAGG-TTGGTTAGTTGGTTGACT 6678

Qy        197 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 256
              | | | |     |  ||   ||||  | | ||| ||    || |  | | | |   | ||
Db       6679 GGTAGGTTGATAGGTAGGTTGGCTAGTTGATTGGTAAGTTGGTTCGTAGGTTGGGAGGTT 6738

Qy        257 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 316
                ||| |   | |  || |||||||| | |   |  || ||  |   | | | |  |  ||
Db       6739 AGTTGGTTAGTTGGTTGGTTGCTTGGTTGGCAGGTTGGTTTGTATGTTGGTAGGTTAGTT 6798

Qy        317 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 376
              | ||| | | | |  || | | |||   | | |   | ||  ||| |   | |  || |
Db       6799 GATTGGTTGGTAGGTTGGTAGATTGGCAGGTTGGTGGGTTAGTTGGTTAGTTGGTTGGTA 6858

Qy        377 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 436
              | ||| | |      || | | ||| | | |||  || ||| | ||  || ||   || |
Db       6859 GGTTGGTTGGCAAGTTGATAGGTTGGTTGGTGGGTTGGTTGATAAGTTGATAGGTTGGTT 6918

Qy        437 GCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCT 496
              | |   ||| || | |  |   |  | ||  ||||   |||||||   || || | || |
Db       6919 GGTAAGTTGATAGGTTGGTTGGTGGGTTGGTTGATAAGTTGATAGGTTGGTTGGTAGGTT 6978

Qy        497 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG----GCTGCTAGTAGGGCTGCTG 552
              | | ||  || || | |  ||  ||   || || | |    | || |||| || || |
Db       6979 GGTTGTTAGGGTGGTAGATTGGCAGGTTGGTTGGTAGTTTTGTTGGTAGTTTGGTTGGTT 7038

Qy        553 GGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTG 612
               |||| | |   || |  || | || | |   |   | | ||  | | |  || |  |
Db       7039 AGCTGGTTGGTTGGTTAGTGAGTTGGTTGGTAGATAGATTGGTAGATTGGTTGGTTGGTT 7098

Qy        613 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 672
               | |  | |   || |  | || |  |     || |  | |  || |||   || || |
Db       7099 AGTTTGTTGGTAGGTTAGTAGGTTTGTTAGTTGGTTAGTTGATTGGTAGGTTGGTTGGTT 7158

Qy        673 GGCTGCTAGTAGGGCTGCCT 692
              | |||  || |   || |||
Db       7159 GACTGAAAGGACCCCTTCCT 7178


RESULT 14
AAK80024
ID   AAK80024 standard; DNA; 14212 BP.
XX
AC   AAK80024;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34836.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
```

PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 34836; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 14212 BP; 3809 A; 3421 C; 3440 G; 3542 T; 0 other;

Query Match 10.0%; Score 99.6; DB 22; Length 14212;
Best Local Similarity 51.1%; Pred. No. 1.8e-15;
Matches 286; Conservative 0; Mismatches 269; Indels 5; Gaps 2;

Qy 137 GCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 196
Db 6620 GTTGACTGGTTAGTAGGTTGGTTAGTTGGTTGGCTGGTAGG-TTGGTTAGTTGGTTGACT 6678

Qy 197 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 256
Db 6679 GGTAGGTTGATAGGTAGGTTGGCTAGTTGATTGGTAAGTTGGTTCGTAGGTTGGGAGGTT 6738

Qy 257 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 316
Db 6739 AGTTGGTTAGTTGGTTGGTTGCTTGGTTGGCAGGTTGGTTTGTATGTTGGTAGGTTAGTT 6798

Qy 317 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 376
Db 6799 GATTGGTTGGTAGGTTGGTAGATTGGCAGGTTGGTGGGTTAGTTGGTTAGTTGGTTGGTA 6858

Qy 377 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 436
Db 6859 GGTTGGTTGGCAAGTTGATAGGTTGGTTGGTGGGTTGGTTGATAAGTTGATAGGTTGGTT 6918

Qy 437 GCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCT 496
Db 6919 GGTAAGTTGATAGGTTGGTTGGTGGGTTGGTTGATAAGTTGATAGGTTGGTTGGTAGGTT 6978

```
Qy       497 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG----GCTGCTAGTAGGGCTGCTG 552
             | | ||  || || | |  ||  ||   || || | |    | || ||||  || || |
Db      6979 GGTTGTTAGGGTGGTAGATTGGCAGGTTGGTTGGTAGTTTTGTTGGTAGTTTGGTTGGTT 7038

Qy       553 GGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTG 612
              |||| | |   || |  || | || | |   |   | | ||  | | |   || |  |
Db      7039 AGCTGGTTGGTTGGTTAGTGAGTTGGTTGGTAGATAGATTGGTAGATTGGTTGGTTGGTT 7098

Qy       613 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 672
             | |  | |   || |  | || |  |     || |  | |  || |||   || || |
Db      7099 AGTTTGTTGGTAGGTTAGTAGGTTTGTTAGTTGGTTAGTTGATTGGTAGGTTGGTTGGTT 7158

Qy       673 GGCTGCTAGTAGGGCTGCCT 692
             | |||  || |   || |||
Db      7159 GACTGAAAGGACCCCTTCCT 7178


RESULT 15
AAK85921
ID   AAK85921 standard; DNA; 14212 BP.
XX
AC   AAK85921;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40733.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
```

```
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 40733; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 14212 BP; 3809 A; 3421 C; 3440 G; 3542 T; 0 other;

  Query Match             10.0%;  Score 99.6;  DB 22;  Length 14212;
  Best Local Similarity   51.1%;  Pred. No. 1.8e-15;
  Matches  286;  Conservative    0;  Mismatches  269;  Indels    5;  Gaps    2;

Qy        137 GCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 196
              | |||  ||  ||||||   | | | ||   ||   |||||  || ||  | | | | ||
Db       6620 GTTGACTGGTTAGTAGGTTGGTTAGTTGGTTGGCTGGTAGG-TTGGTTAGTTGGTTGACT 6678

Qy        197 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 256
              | | | |     |  ||   ||||  | | ||| ||    || |  | | | |   | ||
Db       6679 GGTAGGTTGATAGGTAGGTTGGCTAGTTGATTGGTAAGTTGGTTCGTAGGTTGGGAGGTT 6738

Qy        257 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 316
              |||  |   | |  || |||||| | |   |  || ||  |   | | | |  ||
Db       6739 AGTTGGTTAGTTGGTTGGTTGCTTGGTTGGCAGGTTGGTTTGTATGTTGGTAGGTTAGTT 6798

Qy        317 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 376
              | ||| | | | |  || | | |||   | | |   | ||  ||| |   | |  || |
Db       6799 GATTGGTTGGTAGGTTGGTAGATTGGCAGGTTGGTGGGTTAGTTGGTTAGTTGGTTGGTA 6858

Qy        377 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 436
              | ||| | |       || | | ||| | | |||  || ||| | ||  || ||   || |
Db       6859 GGTTGGTTGGCAAGTTGATAGGTTGGTTGGTGGGTTGGTTGATAAGTTGATAGGTTGGTT 6918

Qy        437 GCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCT 496
              | |   ||| || | |   |   |  ||  ||||   |||||||   || || | || |
Db       6919 GGTAAGTTGATAGGTTGGTTGGTGGGTTGGTTGATAAGTTGATAGGTTGGTTGGTAGGTT 6978

Qy        497 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG----GCTGCTAGTAGGGCTGCTG 552
              | | ||  || || | |  ||  ||   || || | |    | || ||||  || || |
Db       6979 GGTTGTTAGGGTGGTAGATTGGCAGGTTGGTTGGTAGTTTTGTTGGTAGTTTGGTTGGTT 7038

Qy        553 GGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTG 612
              ||||  | |   || |  || | || | | |   |  | | ||  | | |   || |  |
Db       7039 AGCTGGTTGGTTGGTTAGTGAGTTGGTTGGTAGATAGATTGGTAGATTGGTTGGTTGGTT 7098

Qy        613 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 672
              | |  | |   || |  | || | |     || |  | |  || ||| |   || || |
Db       7099 AGTTTGTTGGTAGGTTAGTAGGTTTGTTAGTTGGTTAGTTGATTGGTAGGTTGGTTGGTT 7158

Qy        673 GGCTGCTAGTAGGGCTGCCT 692
              | |||  || |   || |||
Db       7159 GACTGAAAGGACCCCTTCCT 7178


RESULT 16
AAQ95095/c
ID   AAQ95095 standard; DNA; 385 BP.
XX
AC   AAQ95095;
XX
DT   02-OCT-1995  (first entry)
XX
DE   Configuration detecting nucleic acid method target sequence.
XX
KW   Probe; PCR; terminus; amplify; primer; configuration; ss.
XX
OS   Synthetic.
XX
PN   JP06343498-A.
XX
PD   20-DEC-1994.
XX
PF   03-JUN-1993;   93JP-0133640.
XX
PR   03-JUN-1993;   93JP-0133640.
XX
PA   (CANO ) CANON KK.
XX
DR   WPI; 1995-069322/10.
XX
PT   Nucleic acid probe - and method for detecting nucleic acid
XX
PS   Example 1; Column 5; 6pp; Japanese.
XX
CC   The sequence of the PCR amplified target region detected by probes
CC   (AAQ95091-4).  The probes used in a method for the detection of a
CC   nucleic acid target sequence which has PCR-controlled termini and can be
CC   amplified by PCR.  The probes can be used to detect configuration.
XX
SQ   Sequence 385 BP; 93 A; 97 C; 98 G; 97 T; 0 other;

  Query Match              9.9%;  Score 99.2;  DB 16;  Length 385;
  Best Local Similarity   55.5%;  Pred. No. 7.7e-16;
  Matches  191;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;
```

```
Qy         57 GACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 116
              |||    |  |||  ||| ||| |||  ||  || ||| |||       |   | | || |
Db        381 GACGAGCGTGACGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCAT 322

Qy        117 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 176
              ||  || |||  || ||| |||      |   | | || ||| |||      |   | |
Db        321 GCGCGCGCGCGCGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCAT 262

Qy        177 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 236
              || ||| |||  ||  || ||| |||      |   | | |  |||    |   ||
Db        261 GCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATTATATATATATATA 202

Qy        237 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 296
              |  |||  ||  || ||| ||| |||  ||  || ||| ||| |||  ||  || ||| |
Db        201 GCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCAT 142

Qy        297 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 356
              || |||  ||  || ||| ||| |||  ||  || |   | ||| |||    || ||| |
Db        141 GCATGCATGCATGCATGCATGCATGCATGCATAACTACTTACTCTAGCGCATGCAT 82

Qy        357 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
              || |||  ||  || ||| ||| |||  ||  || ||| |  ||
Db         81 GCATGCATGCATGCATGCATGCATGCATGCATGCATCATT 38
```

RESULT 17
AAQ84040/c
ID AAQ84040 standard; DNA; 344 BP.
XX
AC AAQ84040;
XX
DT 03-OCT-1995 (first entry)
XX
DE Plasmid pCU19.
XX
KW Plasmid pUC19; DNA detection metod; ss.
XX
OS Synthetic.
XX
PN JP06343499-A.
XX
PD 20-DEC-1994.
XX
PF 04-JUN-1993; 93JP-0134615.
XX
PR 04-JUN-1993; 93JP-0134615.
XX
PA (CANO) CANON KK.
XX
DR WPI; 1995-069323/10.
XX
PT Method for detecting nucleic acid - using nucleic acid probe
PT and PCR amplification
XX
PS Example 1; Page 4; 9pp; Japanese.
XX
CC AAQ84035 is the plasmid pUC19, used to demonstrate a new method for
CC the detection of DNA.
XX
SQ Sequence 344 BP; 84 A; 85 C; 86 G; 89 T; 0 other;

Query Match 9.7%; Score 96.8; DB 16; Length 344;
Best Local Similarity 55.3%; Pred. No. 3e-15;
Matches 188; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

```
Qy         57 GACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 116
              |||    |  |||  ||| ||| |||  ||  || ||| |||       |   | | || |
Db        340 GACGAGCGTGACGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCAT 281

Qy        117 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 176
              ||  || |||  || ||| |||      |   | | || ||| |||      |   | |
Db        280 GCGCGCGCGCGCGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCAT 221

Qy        177 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 236
              || ||| |||  ||  || ||| |||      |   | | |  |||    |   ||
Db        220 GCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATTATATATATATATA 161

Qy        237 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 296
              |  |||  ||  || ||| ||| |||  |   || ||| ||| |||  ||  || ||| |
Db        160 GCATGCATGCATGCATGCATGCATGCATGGATGCATGCATGCATGCATGCATGCAT 101

Qy        297 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 356
              || |||  ||  || ||| ||| |||  ||  || |   | ||| |||    || ||| |
Db        100 GCATGCATGCATGCATGCATGCATGCATGCATAACTACTTACTCTAGCGCATGCAT 41

Qy        357 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 396
              || |||  ||  || ||| ||| |||  ||  || ||| |
Db         40 GCATGCATGCATGCATGCATGCATGCATGCATGCAT 1
```

RESULT 18
ABZ39851/c
ID ABZ39851 standard; DNA; 2130 BP.
XX
AC ABZ39851;
XX
DT 07-MAR-2003 (first entry)
XX
DE N. gonorrhoeae nucleotide sequence SEQ ID 4291.
XX
KW Antibacterial; infection; vaccine; gene therapy; gene; ds.
XX
OS Neisseria gonorrhoeae.
XX
PN WO200279243-A2.
XX
PD 10-OCT-2002.
XX
PF 12-FEB-2002; 2002WO-IB02069.
XX
PR 12-FEB-2001; 2001GB-0003424.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Fontana MR, Pizza M, Masignani V, Monaci E;
XX
DR WPI; 2003-058415/05.
DR P-PSDB; ABP78881.
XX
PT New protein from Neisseria gonorrheae, useful for the manufacture of a
PT medicament for treating or preventing N. gonorrheae infection -
XX
PS Disclosure; Page 502; 815pp; English.
XX
CC The present invention relates to proteins from Neisseria gonorrhoeae.
CC Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing N. gonorrhoeae
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid
CC molecules of the invention.
XX
SQ Sequence 2130 BP; 792 A; 394 C; 401 G; 543 T; 0 other;

Query Match 9.2%; Score 91.6; DB 25; Length 2130;
Best Local Similarity 71.2%; Pred. No. 1.1e-13;
Matches 121; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

```
Qy        226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
              || | || |  ||    |  ||||  ||| ||| ||| ||||||  ||| ||| ||| |
Db        443 CTGGTTAATTTGGTGTATAATTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 384

Qy        286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
              |||||  ||| ||| ||| ||||||  ||| ||| ||| ||||||  ||| ||| ||| |
```

```
Db        383 CTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 324

Qy        346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 395
              |||||  ||| ||| ||| ||||||  ||| ||| ||| ||||||  |||
Db        323 CTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 274


RESULT 19
ABA15389/c
ID   ABA15389 standard; DNA; 2855 BP.
XX
AC   ABA15389;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 7720.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   20-OCT-2000; 2000US-0242221.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
```

```
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250391.
PR   01-DEC-2000; 2000US-0251160.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system
PT   cancers and metastases -
XX
PS   Disclosure; SEQ ID NO 7720; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2855 BP; 592 A; 829 C; 944 G; 490 T; 0 other;

  Query Match             9.0%;  Score 90;  DB 22;  Length 2855;
  Best Local Similarity   54.5%;  Pred. No. 3.1e-13;
  Matches  180;  Conservative    0;  Mismatches  150;  Indels    0;  Gaps    0;

Qy         75 CTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGC 134
              || |   ||| || | |  | |||   ||   | ||  |     ||||  |||   ||||
Db       1938 CTCCTGACTCCCTTGTCCACGTGCAAGGGCACCTGTGAGTGCATTTGCATGCTTATCTGC 1879

Qy        135 TTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGC 194
               ||||  |   |   |   |  ||| | ||||   |   |   |  ||| |||||     |
Db       1878 ATGCTCATTTGTATGCTTGTTTGCCTACTGATCTGCATGCTCGTTTGCATGCTCATCTAC 1819

Qy        195 CTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGC 254
               ||||  |  |   |  |    |  |  ||||  | ||||  ||     ||||  |||  |  |||
Db       1818 TTGCTCGTTTGCGTGGTCATCTGCGTGCTCGTTTGCGAGCTCATCTGCGTGCTCGTTTGC 1759

Qy        255 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 314
               ||||  |   ||  ||  |  ||  |||||    ||  ||||  |  ||   |||   ||||
Db       1758 GTGCTCGTCTGCGTGCTCGTCTGTGTGCTCATCTGCGTGCTCGTCTGTGTGCTCATCTGC 1699

Qy        315 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 374
               ||||      ||  || || |||| |||||    |  ||||||  |||   |||  | ||||
Db       1698 ATGCTCATCTGCGTGCTTGTTTGCATGCTCATCTGTGTGCTTGTTTGTGTGCTCGTCTGC 1639

Qy        375 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 404
               ||||||  |  ||  ||    ||||  ||  ||
Db       1638 GTGCTTGTTTGCATGCTCATTTGCATGGTC 1609


RESULT 20
AAK89506/c
ID   AAK89506 standard; DNA; 15682 BP.
XX
AC   AAK89506;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human digestive system antigen genomic sequence SEQ ID NO: 3082.
XX
KW   Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW   ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW   digestive system disorder; Meckel's diverticulum; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155314-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01324.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
```

PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases -
XX
PS Disclosure; SEQ ID NO 3082; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention.
XX
SQ Sequence 15682 BP; 3797 A; 4055 C; 4103 G; 3727 T; 0 other;

Query Match 8.4%; Score 84.4; DB 22; Length 15682;
Best Local Similarity 50.5%; Pred. No. 1.4e-11;
Matches 205; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

```
Qy        71 CTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGC 130
             ||||||||||| |   |  |||||| | |      | | |    ||| | |||  ||  ||
Db     12185 CTGCCTGCCTTCCTTCCTGCCTGCCTTCCTTCCTGCCTGCCTTCCTTCCTGCCTGCCTGC 12126

Qy       131 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGC 190
             || | | |      | |      ||||| |||        |     ||||| | |   |
Db     12125 CTTCCTTCCTGCCTGCCTTCCTTCCTGCCTGCCTTCCTTCCTTCCTGCCTGCCTTCCTTC 12066

Qy       191 TGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGC 250
              ||||| |||      | |      | || |   | | |   |   | ||||   |  |
Db     12065 CTGCCTGCCTGCCTTCCTGCCTTCCTGCCTTCATTCATTCCTGCCTGCCTGCCTTCATTC 12006

Qy       251 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 310
             | | ||| |||  ||   ||| | | | ||| |||   ||| | | | ||| |||   |
Db     12005 ATTCCTGCCTGCCTGCCTTCCTTCCTTCCTGCCCGCCTTCCTTCCTTCCTGCCCGCCTTC 11946

Qy       311 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 370
             || | |||  ||   |  ||| | ||| |||   |  |||| | ||| |||  ||  |
Db     11945 CTTCCTGCCCGCCTTCCTTCCTTCCTGCCTGCCTTCCTGCCTTCCTGCCTGCCTGCCTTC 11886

Qy       371 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 430
```

```
Db      11885 ATTCATTCCTGCCTGCCTGCCTTCATTCATTCCTGCCTGCCTTCCTTCATTCCTTCCTTC 11826

Qy        431 AGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTT 476
              |||| ||| |||      | | |   |||| |||     |||
Db      11825 CTTCCTGCCTGCCTGCCTTCCTGCCTGCCTGCCTTCCTTCCTTCTT 11780


RESULT 21
ABS65032
ID   ABS65032 standard; DNA; 42999 BP.
XX
AC   ABS65032;
XX
DT   15-NOV-2002  (first entry)
XX
DE   Invertebrate foraging behaviour associated human DNA sequence #12.
XX
KW   Attention deficit hyperactivity disorder; ADHD; hypertension;
KW   invertebrate foraging behaviour; nitric oxide; hypotensive;
KW   cGMP-dependent protein kinase; human; neuroleptic; ds;
KW   gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200259370-A2.
XX
PD   01-AUG-2002.
XX
PF   13-DEC-2001; 2001WO-US48087.
XX
PR   15-DEC-2000; 2000US-0738630.
XX
PA   (NEUR-) NEUROSCIENCES RES FOUND INC.
XX
PI   Greenspan RJ,  Shaw PJ;
XX
DR   WPI; 2002-636544/68.
XX
PT   Identifying a compound that modulates an attention deficit
PT   hyperactivity disorder (ADHD) for treating e.g. ADHD or hypertension,
PT   comprises measuring a foraging behaviour an invertebrate -
XX
PS   Claim 89; Page 182-195; 246pp; English.
XX
CC   The present invention relates to a method of identifying a compound
CC   that modulates attention deficit hyperactivity disorder (ADHD) in
CC   a mammal. The method comprises administering a test compound to
CC   an invertebrate, and measuring a foraging behaviour of the
CC   invertebrate, where the compound that modulates the foraging behaviour
CC   of the invertebrate is characterised as a compound that modulates
CC   ADHD in a mammal. The compound identified by the new method of the
CC   invention is useful for diagnosing or treating ADHD, hypertension or
CC   other diseases associated with a nitric oxide/cGMP-dependent protein
CC   kinase network in a mammal, particularly humans. The method
CC   distinguishes a compound that has a specific effect on ADHD,
CC   hypertension or other diseases associated with a
CC   nitric oxide/cGMP-dependent kinase protein network
CC   in a mammal from a compound that has a non-specific effect. The
CC   present sequence represents a DNA sequence associated with
CC   invertebrate foraging behaviour.
XX
SQ   Sequence 42999 BP; 6411 A; 13605 C; 11491 G; 11479 T; 13 other;

  Query Match             8.1%;  Score 80.8;  DB 24;  Length 42999;
  Best Local Similarity   58.1%;  Pred. No. 1.6e-10;
  Matches  179;  Conservative    0;  Mismatches  127;  Indels    2;  Gaps    2;

Qy        114 CCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAG 173
              | | |||||| |||  | ||||||||         |  | | || ||   ||
Db      40275 CTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTGTCTTGTTTCTCGATTTCTTTC 40334

Qy        174 TAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAG 233
              |     |  || ||  |  || |||||  ||     ||  |      ||||  ||||||
Db      40335 TTTCTTTTGTTTCTTTCCTGCTTGCTTTCTTGCTTGATTGCTTTCGTGCTTTCTTGCTTT 40394

Qy        234 TAGGGCTGCTCGCTGGC-CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 292
                | || ||  |  |  || ||| ||  || || | |||| || || |||  | |
Db      40395 CTTGTTTTCTTTCTTTCTTTTGTTTCTTTCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTT 40454

Qy        293 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC-TTGCTTGCTCGCTGGCC 351
              ||||||||||  |  || | |||| || || |||  | | | ||  || ||  || ||
Db      40455 GCTTGCTTGCTTTCGTGCTTTCTTGTTTTCTTGCTTTCTTTCTTTTGTTTCTTTCTTGCT 40514

Qy        352 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              |||||  ||||||  || |  | ||||||| || ||| || ||||| | ||||  ||  |
Db      40515 TGCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTTCTTTCT 40574

Qy        412 TGCTTGCT 419
              |||||  ||
Db      40575 TGCTTTCT 40582


RESULT 22
ABX11086
ID   ABX11086 standard; DNA; 42999 BP.
XX
AC   ABX11086;
XX
DT   24-APR-2003  (first entry)
XX
DE   Human ribosomal RNA (rRNA) gene.
XX
KW   Nucleic acid amplification; artificial chromosome isolation; MAC;
KW   DNA delivery; mammalian artificial chromosome; gene therapy; organ;
KW   humanised genetically transformed animal; chromosomal element;
KW   gene product production system; transgenic; centromere function;
KW   information storage vehicle; artificial chromosome vector; human;
KW   species-specific artificial chromosome; ribosomal RNA; rRNA; gene; ds.
XX
OS   Homo sapiens.
XX
PN   US2002160410-A1.
XX
PD   31-OCT-2002.
XX
PF   17-APR-2002; 2002US-0125767.
XX
PR   28-NOV-2000; 2000US-0724693.
PR   10-APR-1997;   97US-0835682.
PR   10-APR-1996;   96US-0629822.
PR   15-JUL-1996;   96US-0682080.
PR   07-AUG-1996;   96US-0695191.
XX
PA   (HADL/) HADLACZKY G.
PA   (SZAL/) SZALAY A A.
XX
PI   Hadlaczky G,  Szalay AA;
XX
DR   WPI; 2003-265757/26.
XX
PT   Amplifying nucleic acid for constructing artificial chromosomes,
PT   comprises introducing a DNA fragment into a cell to incorporate the DNA
PT   into a chromosome, and identifying cells with chromosomes that has
PT   undergone amplification  -
XX
PS   Disclosure; Page 76-94; 107pp; English.
XX
CC   The present invention relates to methods for amplifying nucleic acids
CC   in cells, for isolating artificial chromosomes, and preparing cell
CC   lines that contain artificial chromosomes. The methods comprise
CC   introducing a DNA fragment comprising a selectable marker into a cell,
CC   growing the cell under selective conditions to produce cells that have
CC   incorporated the DNA fragment or its portion into a chromosome, and
CC   identifying from among the resulting cells those that include a
CC   chromosome or its fragment with a portion that has undergone
```

```
CC   amplification. The invention also discloses cells and cell lines that
CC   contain the nucleic acids or artificial chromosomes. The methods of
CC   the invention are useful for amplifying nucleic acids in cells, in
CC   generating and isolating artificial chromosomes (e.g. mammalian
CC   artificial chromosomes or MACs), and in delivering the chromosomes
CC   to selected cells and tissues. The artificial chromosomes are useful
CC   in gene therapy, gene product production systems, production
CC   of humanised genetically transformed animal organs, production of
CC   transgenic plants and animals that would employ chromosomal elements as
CC   information storage vehicles, for analysis and study of centromere
CC   function, for the production of artificial chromosome vectors, and for
CC   the preparation of species-specific artificial chromosomes. The present
CC   sequence representing the human ribosomal RNA (rRNA) gene is used in
CC   the method of the present invention.
CC   Note: All N nucleotides (EXCEPT those at positions 21526-21527, 23183,
CC   24975-24977, 28699, 28945, 34074 and 34474) have been added to the
CC   sequence by the indexer to match the number of bases this sequence
CC   is listed to contain in the sequence listing.
XX
SQ   Sequence 42999 BP; 5970 A; 12784 C; 10787 G; 10765 T; 2693 other;

  Query Match             8.1%;  Score 80.8;  DB 25;  Length 42999;
  Best Local Similarity   49.4%;  Pred. No. 1.6e-10;
  Matches  265;  Conservative    0;  Mismatches  263;  Indels    8;  Gaps    4;

Qy        182 CTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTG 241
              || ||||||| || ||||||||   |         |  | |  | || || |      |
Db      39871 CTCGCTCGCTTGCTTGCTTGCTTTCGNNNNTTCTTGCTTTCCCGTTTTCTTGCTTTCTTT 39930

Qy        242 CTCGCTGGC-CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 300
              ||  ||  |  | ||| | |||| ||     ||||||||||| |||  | ||||| |||
Db      39931 CTTTCTTTCGTTTCTTTCATGCTTGCNNNNTTGCTTGCTTGCTTGCTTTCGTGCTTTCTT 39990

Qy        301 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC-TTGCTCGCTGGCCTGCTTGCT 359
              |||  |  |  | ||| ||| ||  |     | ||| | ||  |  ||  | ||||||||
Db      39991 GCTTTCCTGTTTTCTTTCTTTCTTTCNNNNTTTCTTTCTTTTGTTTCTTTCTTGCTTGCT 40050

Qy        360 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              | || ||| || |||||||||| |  |    | ||||| | || |  ||  | | ||| ||
Db      40051 TTCTTGCTTGCTTGCTTGCTTTCGTGNNNNCTTGCTTTCCTGTTTTCTTTCTTTCTTTCT 40110

Qy        420 GAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGAT 479
              |  |     | ||||||||||     |   || | | || |  |||| |
Db      40111 TTCTTTTCTTT---CTTTCTTGCTTGCTTNNNNGCTTGCTTGCTTTCGTGCTTTCTTGTT 40167

Qy        480 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 539
              |  | ||    | ||  |   | | ||    |||| |      |||| |  ||||
Db      40168 TTCTCGATTTCTTTCTTTCTTTT---GTTTCTNNNNTGCTTGCTTTCTTGCTTGCTTGCT 40224

Qy        540 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 599
              |  |  ||   | ||  |     || |    |  |  |   | | ||    ||||
Db      40225 TTCGTGCTTCTTGCTTTCCTGTTTTCTTTCTTNNNNTCTTTCTTTTGTTTCTTTCTTGCT 40284

Qy        600 AGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 659
              |      | ||  ||||     |    ||    |      |    | |    |  |
Db      40285 TGCTTTCTTGCTTGCTTGCTTTCGTGCTGTCTNNNNTCTCGATTTCTTTCTTTCTTTTGT 40344

Qy        660 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCT 715
              |||||  |  | ||  |   |  |||      |||| ||||||| |||| |
Db      40345 TTCTTTCCTGCTTGCTTTCTTGCTTGATTGCTNNNNTGCTTTCTTGCTTTCTTGTT 40400


RESULT 23
AAT42063
ID   AAT42063 standard; DNA; 1830121 BP.
XX
AC   AAT42063;
XX
DT   14-SEP-1999  (first entry)
XX
DE   Haemophilus influenzae complete genome sequence.
XX
KW   Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW   expression modulating fragment; regulation; gene expression; vector;
KW   organism; open reading frame; ORF; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO9633276-A1.
XX
PD   24-OCT-1996.
XX
PF   22-APR-1996;   96WO-US05320.
XX
PR   07-JUN-1995;   95US-0487429.
PR   21-APR-1995;   95US-0426787.
PR   07-JUN-1995;   95US-0476102.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (UYJO ) UNIV JOHNS HOPKINS.
XX
PI   Adams MD,  Fleischmann RD,  Smith HO,  Venter JC,  White O;
XX
DR   WPI; 1996-485782/48.
XX
PT   Haemophilus influenzae Rd genome recorded on computer readable
PT   medium  - useful for identifying commercially important nucleic acid
PT   fragments by homology searching
XX
PS   Claim 1; Page 77.2-77.1091; 1291pp; English.
XX
CC   This sequence represents the complete genome sequence of the bacterium
CC   Haemophilus influenzae strain Rd.  The invention relates to a computer
CC   readable medium (CRM) having recorded upon it the complete H.influenzae
CC   nucleotide sequence (I), a representative fragment of (I) or a nucleotide
CC   sequence at least 99% identical to (I).  By providing the full-length
CC   genomic sequence in a computer readable form, it is possible to identify
CC   commercially important nucleic acid fragments and expression modulating
CC   fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC   regulate the expression of a nucleic acid molecule.  Vectors and altered
CC   organisms comprising the predicted ORFS can be used to produce any of the
CC   polypeptide fragments of the H. influenzae Rd genome.
XX
SQ   Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

  Query Match             8.0%;  Score 79.8;  DB 17;  Length 1830121;
  Best Local Similarity   84.1%;  Pred. No. 8.5e-10;
  Matches   90;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy        255 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 314
              ||||||||| ||| || |||||||||||| ||| || |||||||||||| ||| || |||
Db    1543129 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 1543188

Qy        315 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 361
              ||||||||| ||| || |||||||||||| ||| || |||||  |||
Db    1543189 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTAATTG 1543235


RESULT 24
ABQ31994
ID   ABQ31994 standard; DNA; 1172 BP.
XX
AC   ABQ31994;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 18585.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
```

```
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 1172 BP; 3 A; 583 C; 583 G; 3 T; 0 other;

  Query Match             7.9%;  Score 79;  DB 24;  Length 1172;
  Best Local Similarity   48.2%;  Pred. No. 1.5e-10;
  Matches  223;  Conservative    0;  Mismatches  240;  Indels    0;  Gaps    0;

Qy         12 CGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGC 71
              |||| | | |   | |  |  | |   ||| | ||| ||  | | |    || |  ||
Db         37 CGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTACGCGCGCGCGCG 96

Qy         72 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 131
              ||  ||  || ||| ||  ||  ||    | |   |   ||  ||  || ||| ||
Db         97 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 156

Qy        132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191
              ||  ||    | |   |  ||  ||  ||    | |   ||| ||  ||  || |||
Db        157 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCGCG 216

Qy        192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251
              ||  ||  ||    | |   |   |  ||  ||  ||  |   |   ||  ||| ||
Db        217 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 276

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
              ||  ||  || ||| ||  ||  ||  || ||| ||  ||  ||  ||  || ||| ||
Db        277 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 336

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
              ||  ||  || |||  ||  ||  ||  || |||  ||  ||  ||  || |||  ||
Db        337 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 396

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTA 431
              ||  ||  || |||  ||  ||  ||  || |||  ||  ||  ||    | |   |

Db        397 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 456

Qy        432 GGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGC 474
                |   ||  ||  ||    | |   |  |||  ||  |   ||
Db        457 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 499


RESULT 25
ABQ31995/c
ID   ABQ31995 standard; DNA; 1172 BP.
XX
AC   ABQ31995;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 18586.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 1172 BP; 3 A; 583 C; 583 G; 3 T; 0 other;

  Query Match             7.9%;  Score 79;  DB 24;  Length 1172;
  Best Local Similarity   48.2%;  Pred. No. 1.5e-10;
  Matches  223;  Conservative    0;  Mismatches  240;  Indels    0;  Gaps    0;

Qy         12 CGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGC 71
```

```
Db       1136 CGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTACGCGCGCGCGCG 1077

Qy         72 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 131
Db       1076 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1017

Qy        132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191
Db       1016 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCG 957

Qy        192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251
Db        956 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 897

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
Db        896 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 837

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
Db        836 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 777

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTA 431
Db        776 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 717

Qy        432 GGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGC 474
Db        716 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 674


RESULT 26
ABQ31996
ID   ABQ31996 standard; DNA; 1172 BP.
XX
AC   ABQ31996;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 18587.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 1172 BP; 0 A; 583 C; 586 G; 3 T; 0 other;

  Query Match             7.9%;  Score 79;  DB 24;  Length 1172;
  Best Local Similarity   48.2%;  Pred. No. 1.5e-10;
  Matches  223;  Conservative    0;  Mismatches  240;  Indels    0;  Gaps    0;

Qy         12 CGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGC 71
Db        709 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 768

Qy         72 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 131
Db        769 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 828

Qy        132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191
Db        829 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 888

Qy        192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251
Db        889 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 948

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
Db        949 CGCGCGCGCGCGCGCGCGCGCGCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1008

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
Db       1009 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1068

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTA 431
Db       1069 CGCGCGCGCGCGCGCGCGCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTG 1128

Qy        432 GGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGC 474
Db       1129 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1171


RESULT 27
ABQ31997/c
ID   ABQ31997 standard; DNA; 1172 BP.
XX
AC   ABQ31997;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 18588.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
```

```
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 1172 BP; 3 A; 586 C; 583 G; 0 U; 0 other;

  Query Match             7.9%;  Score 79;  DB 24;  Length 1172;
  Best Local Similarity   48.2%;  Pred. No. 1.5e-10;
  Matches  223;  Conservative    0;  Mismatches  240;  Indels    0;  Gaps    0;

Qy         12 CGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGC 71
              |||| | |     | |  |  | |   ||| | ||| ||  | | |    || |  ||
Db        464 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 405

Qy         72 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 131
                ||  ||  || |||  ||  ||  ||    | |   |   ||  ||  || |||  ||
Db        404 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 345

Qy        132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191
                ||  ||    | |   |   ||  ||  ||    | |   |   ||  ||  || |||
Db        344 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 285

Qy        192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251
                ||  ||  ||    | |   |   |   ||  ||  ||  |   |   || |||  ||
Db        284 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 225

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
                ||  ||  || |||  ||  || |||  || |||  |  ||  ||  || |||  ||
Db        224 CGCGCGCGCGCGCGCGCGCGCGCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 165

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
                ||  ||  || |||  ||  ||  ||  || |||  ||  ||  ||  || |||  ||
Db        164 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTA 431
                ||  ||  || |||  || |||  ||  || |||  ||  ||  ||    | |   ||
Db        104 CGCGCGCGCGCGCGCGCGCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTG 45

Qy        432 GGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGC 474
                |   ||  ||  ||    | |   |  |||  ||  |   ||
Db         44 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 2


RESULT 28
ABQ97027/c
ID   ABQ97027 standard; cDNA; 388 BP.
XX
AC   ABQ97027;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Mouse ES cell related cDNA SEQ ID NO 295.
XX
KW   Mouse; ES cell; gene trapped sequence; GTS; gene expression;
KW   development disorder; cell differentiation disorder; gene; ss.
XX
OS   Mus sp.
XX
PN   US2002081668-A1.
XX
PD   27-JUN-2002.
XX
PF   30-NOV-2000; 2000US-0728446.
XX
PR   20-NOV-1998;   98US-109302P.
PR   01-DEC-1999;   99US-168270P.
XX
PA   (FRIE/) FRIEDRICH G.
PA   (ZAMB/) ZAMBROWICZ B.
PA   (SAND/) SANDS A T.
XX
PI   Friedrich G,  Zambrowicz B,  Sands AT;
XX
DR   WPI; 2002-626541/67.
XX
PT   Novel murine polynucleotides that individually identify novel genes
PT   into which a retroviral gene trap vector has been integrated, useful in
PT   genomic analysis and in discovery, development of therapeutic and
PT   diagnostic agents  -
XX
PS   Claim 2; SEQ ID NO 295; 29pp + Sequence Listing; English.
XX
CC   The invention relates to isolated murine polynucleotides (I) comprising a
CC   contiguous stretch of at least about 60 nucleotides of a sequence
CC   (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences
CC   (GTSs). The novel genes can be used in a process to identify novel
CC   polynucleotide sequences by comparing them to the novel gene sequences.
CC   The novel genes and cells are useful in functional genomic analysis and
CC   in the discovery and development of new therapeutic and diagnostic agents
CC   and methods. (I) is useful for identifying the coding regions of the
CC   murine genome, to isolate cDNAs, genomic clones or full-length
CC   genes/polynucleotides or homologues, heterologues, paralogues or
CC   orthologues that are capable of hybridising to one or more of the GTSs
CC   under stringent conditions. (I) can be incorporated into a phage display
CC   system that can be used to screen for proteins or other ligands, that are
CC   capable of binding an amino acid sequence encoded by an oligonucleotide
CC   or polynucleotide sequence in at least one of the GTS sequences. (I) is
CC   useful in arrays, such as gene chips, to identify and characterise
CC   temporal and tissue specific gene expression, to identify the gene of
CC   interest from many sources and for genetic manipulations such as
CC   antisense inhibition and gene targeting. Decreasing the level of
CC   expression of (I) and/or down regulating the activity of peptides or
CC   proteins encoded by (I) is useful for treating development and cell
CC   differentiation disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at sequdata.uspto.gov/sequence.html?DocID=20020081668.
```

```
XX
SQ   Sequence 388 BP; 133 A; 140 C; 62 G; 51 T; 2 other;

  Query Match             7.8%;  Score 78.4;  DB 24;  Length 388;
  Best Local Similarity   60.2%;  Pred. No. 1.6e-10;
  Matches  130;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
              | ||||  | |  |     ||||   |   |||||   |  | ||  | | ||| | | |
Db        218 CCTGCTGTTGGTCCCTGCTGCTGCATTTGCTGCTTTTGCTATTGCTGGATGGCTGGATGG 159

Qy        286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
              ||||  |||| | | ||| | ||||  | |||||  | | |||||   ||| | ||| | |
Db        158 CTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGGTGG 99

Qy        346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 405
              ||||  | | ||| | | |||||  |||| | | ||| | ||||  | ||| | | | |
Db         98 ATGGCTGGATGGCTGGATGGCTGGTTTGCTGGATGGCTGGATGGCTGGGTTGGTGGATGG 39

Qy        406 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTG 441
              ||||   ||| | |  | |||  || ||  |||| |
Db         38 CTGGATGGCTGGGTTGGTGGATTGTTGGATTGCTGG 3


RESULT 29
ABQ97101/c
ID   ABQ97101 standard; cDNA; 428 BP.
XX
AC   ABQ97101;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Mouse ES cell related cDNA SEQ ID NO 369.
XX
KW   Mouse; ES cell; gene trapped sequence; GTS; gene expression;
KW   development disorder; cell differentiation disorder; gene; ss.
XX
OS   Mus sp.
XX
PN   US2002081668-A1.
XX
PD   27-JUN-2002.
XX
PF   30-NOV-2000; 2000US-0728446.
XX
PR   20-NOV-1998;   98US-109302P.
PR   01-DEC-1999;   99US-168270P.
XX
PA   (FRIE/) FRIEDRICH G.
PA   (ZAMB/) ZAMBROWICZ B.
PA   (SAND/) SANDS A T.
XX
PI   Friedrich G,  Zambrowicz B,  Sands AT;
XX
DR   WPI; 2002-626541/67.
XX
PT   Novel murine polynucleotides that individually identify novel genes
PT   into which a retroviral gene trap vector has been integrated, useful in
PT   genomic analysis and in discovery, development of therapeutic and
PT   diagnostic agents
XX
PS   Claim 2; SEQ ID NO 369; 29pp + Sequence Listing; English.
XX
CC   The invention relates to isolated murine polynucleotides (I) comprising a
CC   contiguous stretch of at least about 60 nucleotides of a sequence
CC   (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences
CC   (GTSs). The novel genes can be used in a process to identify novel
CC   polynucleotide sequences by comparing them to the novel gene sequences.
CC   The novel genes and cells are useful in functional genomic analysis and
CC   in the discovery and development of new therapeutic and diagnostic agents
CC   and methods. (I) is useful for identifying the coding regions of the
CC   murine genome, to isolate cDNAs, genomic clones or full-length
CC   genes/polynucleotides or homologues, heterologues, paralogues or
CC   orthologues that are capable of hybridising to one or more of the GTSs
CC   under stringent conditions. (I) can be incorporated into a phage display
CC   system that can be used to screen for proteins or other ligands, that are
CC   capable of binding an amino acid sequence encoded by an oligonucleotide
CC   or polynucleotide sequence in at least one of the GTS sequences. (I) is
CC   useful in arrays, such as gene chips, to identify and characterise
CC   temporal and tissue specific gene expression, to identify the gene of
CC   interest from many sources and for genetic manipulations such as
CC   antisense inhibition and gene targeting. Decreasing the level of
CC   expression of (I) and/or down regulating the activity of peptides or
CC   proteins encoded by (I) is useful for treating development and cell
CC   differentiation disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at sequdata.uspto.gov/sequence.html?DocID=20020081668.
XX
SQ   Sequence 428 BP; 149 A; 158 C; 66 G; 55 T; 0 other;

  Query Match             7.8%;  Score 78.4;  DB 24;  Length 428;
  Best Local Similarity   60.2%;  Pred. No. 1.6e-10;
  Matches  130;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
              | ||||  | |  |     ||||   |   |||||   |  | ||  | | ||| | | |
Db        259 CCTGCTGTTGGTCCCTGCTGCTGCATTTGCTGCTTTTGCTATTGCTGGATGGCTGGATGG 200

Qy        286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
              ||||  |||| | | ||| | ||||  | |||||  | | |||||   ||| | ||| | |
Db        199 CTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGGTGG 140

Qy        346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 405
              ||||  | | ||| | | |||||  |||| | | ||| | ||||  | ||| | | | |
Db        139 ATGGCTGGATGGCTGGATGGCTGGTTTGCTGGATGGCTGGATGGCTGGGTTGGTGGATGG 80

Qy        406 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTG 441
              ||||   ||| | |  | |||  || ||  |||| |
Db         79 CTGGATGGCTGGGTTGGTGGATTGTTGGATTGCTGG 44


RESULT 30
ABQ97086/c
ID   ABQ97086 standard; cDNA; 436 BP.
XX
AC   ABQ97086;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Mouse ES cell related cDNA SEQ ID NO 354.
XX
KW   Mouse; ES cell; gene trapped sequence; GTS; gene expression;
KW   development disorder; cell differentiation disorder; gene; ss.
XX
OS   Mus sp.
XX
PN   US2002081668-A1.
XX
PD   27-JUN-2002.
XX
PF   30-NOV-2000; 2000US-0728446.
XX
PR   20-NOV-1998;   98US-109302P.
PR   01-DEC-1999;   99US-168270P.
XX
PA   (FRIE/) FRIEDRICH G.
PA   (ZAMB/) ZAMBROWICZ B.
PA   (SAND/) SANDS A T.
XX
PI   Friedrich G,  Zambrowicz B,  Sands AT;
XX
DR   WPI; 2002-626541/67.
XX
PT   Novel murine polynucleotides that individually identify novel genes
PT   into which a retroviral gene trap vector has been integrated, useful in
```

```
PT   genomic analysis and in discovery, development of therapeutic and
PT   diagnostic agents -
XX
PS   Claim 2; SEQ ID NO 354; 29pp + Sequence Listing; English.
XX
CC   The invention relates to isolated murine polynucleotides (I) comprising a
CC   contiguous stretch of at least about 60 nucleotides of a sequence
CC   (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences
CC   (GTSs). The novel genes can be used in a process to identify novel
CC   polynucleotide sequences by comparing them to the novel gene sequences.
CC   The novel genes and cells are useful in functional genomic analysis and
CC   in the discovery and development of new therapeutic and diagnostic agents
CC   and methods. (I) is useful for identifying the coding regions of the
CC   murine genome, to isolate cDNAs, genomic clones or full-length
CC   genes/polynucleotides or homologues, heterologues, paralogues or
CC   orthologues that are capable of hybridising to one or more of the GTSs
CC   under stringent conditions. (I) can be incorporated into a phage display
CC   system that can be used to screen for proteins or other ligands, that are
CC   capable of binding an amino acid sequence encoded by an oligonucleotide
CC   or polynucleotide sequence in at least one of the GTS sequences. (I) is
CC   useful in arrays, such as gene chips, to identify and characterise
CC   temporal and tissue specific gene expression, to identify the gene of
CC   interest from many sources and for genetic manipulations such as
CC   antisense inhibition and gene targeting. Decreasing the level of
CC   expression of (I) and/or down regulating the activity of peptides or
CC   proteins encoded by (I) is useful for treating development and cell
CC   differentiation disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at sequdata.uspto.gov/sequence.html?DocID=20020081668.
XX
SQ   Sequence 436 BP; 150 A; 157 C; 68 G; 60 T; 1 other;

  Query Match             7.8%;  Score 78.4;  DB 24;  Length 436;
  Best Local Similarity   60.2%;  Pred. No. 1.6e-10;
  Matches  130;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
              | ||||  | |  |     ||||   |   |||||   |  | ||  | | ||| | | |
Db        267 CCTGCTGTTGGTCCCTGCTGCTGCATTTGCTGCTTTTGCTATTGCTGGATGGCTGGATGG 208

Qy        286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
              ||||  |||| | | ||| | ||||  | ||||| | | |||||   ||| | ||| | |
Db        207 CTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGGTGG 148

Qy        346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 405
               ||||  | | ||| | | |||||  |||| | | ||| | ||||  | ||| | | | |
Db        147 ATGGCTGGATGGCTGGATGGCTGGTTTGCTGGATGGCTGGATGGCTGGGTTGGTGGATGG 88

Qy        406 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTG 441
              ||||   ||| | |  | |||  || ||  |||| |
Db         87 CTGGATGGCTGGGTTGGTGGATTGTTGGATTGCTGG 52


RESULT 31
AAL40273/c
ID   AAL40273 standard; DNA; 169 BP.
XX
AC   AAL40273;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 35.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   N-PSDB; AAO21920, AAO21921.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear -
XX
PS   Example 1; Fig 3; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 169 BP; 69 A; 38 C; 33 G; 29 T; 0 other;

  Query Match             7.8%;  Score 78;  DB 24;  Length 169;
  Best Local Similarity   85.3%;  Pred. No. 1.6e-10;
  Matches   87;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy        239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
              ||||| ||| || |||||||||||| ||| || |||||||||||| ||| || |||||||
Db        112 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 53

Qy        299 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 340
              ||||| ||| || |||||||||||| ||| || |||||||||
Db         52 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 11


RESULT 32
ABQ97094/c
ID   ABQ97094 standard; cDNA; 409 BP.
XX
AC   ABQ97094;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Mouse ES cell related cDNA SEQ ID NO 362.
XX
KW   Mouse; ES cell; gene trapped sequence; GTS; gene expression;
KW   development disorder; cell differentiation disorder; gene; ss.
```

XX
OS Mus sp.
XX
PN US2002081668-A1.
XX
PD 27-JUN-2002.
XX
PF 30-NOV-2000; 2000US-0728446.
XX
PR 20-NOV-1998; 98US-109302P.
PR 01-DEC-1999; 99US-168270P.
XX
PA (FRIE/) FRIEDRICH G.
PA (ZAMB/) ZAMBROWICZ B.
PA (SAND/) SANDS A T.
XX
PI Friedrich G, Zambrowicz B, Sands AT;
XX
DR WPI; 2002-626541/67.
XX
PT Novel murine polynucleotides that individually identify novel genes
PT into which a retroviral gene trap vector has been integrated, useful in
PT genomic analysis and in discovery, development of therapeutic and
PT diagnostic agents -
XX
PS Claim 2; SEQ ID NO 362; 29pp + Sequence Listing; English.
XX
CC The invention relates to isolated murine polynucleotides (I) comprising a
CC contiguous stretch of at least about 60 nucleotides of a sequence
CC (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences
CC (GTSs). The novel genes can be used in a process to identify novel
CC polynucleotide sequences by comparing them to the novel gene sequences.
CC The novel genes and cells are useful in functional genomic analysis and
CC in the discovery and development of new therapeutic and diagnostic agents
CC and methods. (I) is useful for identifying the coding regions of the
CC murine genome, to isolate cDNAs, genomic clones or full-length
CC genes/polynucleotides or homologues, heterologues, paralogues or
CC orthologues that are capable of hybridising to one or more of the GTSs
CC under stringent conditions. (I) can be incorporated into a phage display
CC system that can be used to screen for proteins or other ligands, that are
CC capable of binding an amino acid sequence encoded by an oligonucleotide
CC or polynucleotide sequence in at least one of the GTS sequences. (I) is
CC useful in arrays, such as gene chips, to identify and characterise
CC temporal and tissue specific gene expression, to identify the gene of
CC interest from many sources and for genetic manipulations such as
CC antisense inhibition and gene targeting. Decreasing the level of
CC expression of (I) and/or down regulating the activity of peptides or
CC proteins encoded by (I) is useful for treating development and cell
CC differentiation disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at sequdata.uspto.gov/sequence.html?DocID=20020081668.
XX
SQ Sequence 409 BP; 135 A; 153 C; 63 G; 53 T; 5 other;

Query Match 7.7%; Score 77.4; DB 24; Length 409;
Best Local Similarity 59.7%; Pred. No. 2.9e-10;
Matches 129; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
Db 246 CCTGCTGTTGGTCCCTGCTGCTGCATTTGCTGCTTTTGCTATTGCTGGATGGCTGGATGG 187

Qy 286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
Db 186 CTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGGTGG 127

Qy 346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 405
Db 126 ATGGCTGGATGGCTGGATGGCTGGTTTGCTGGATGGCTGGATGGCTGGGTTGGTGGATGG 67

Qy 406 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTG 441

Db 66 CTGGATGGCTGGGTTGGTGGATTGTTGGATTGNTGG 31

RESULT 33
ABQ31994/c
ID ABQ31994 standard; DNA; 1172 BP.
XX
AC ABQ31994;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 18585.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 1172 BP; 3 A; 583 C; 583 G; 3 T; 0 other;

Query Match 7.7%; Score 77.4; DB 24; Length 1172;
Best Local Similarity 47.9%; Pred. No. 4e-10;
Matches 222; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

Qy 12 CGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGC 71
Db 646 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 587

Qy 72 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 131

```
Db        586 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 527

Qy        132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191

Db        526 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 467

Qy        192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251

Db        466 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 407

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311

Db        406 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 347

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371

Db        346 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 287

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTA 431

Db        286 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 227

Qy        432 GGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGC 474

Db        226 CGCGCGCGCGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCGC 184


RESULT 34
ABQ31995
ID   ABQ31995 standard; DNA; 1172 BP.
XX
AC   ABQ31995;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 18586.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 1172 BP; 3 A; 583 C; 583 G; 3 T; 0 other;

  Query Match             7.7%;  Score 77.4;  DB 24;  Length 1172;
  Best Local Similarity   47.9%;  Pred. No. 4e-10;
  Matches  222;  Conservative    0;  Mismatches  241;  Indels    0;  Gaps    0;

Qy         12 CGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGC 71

Db        527 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 586

Qy         72 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 131

Db        587 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 646

Qy        132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191

Db        647 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 706

Qy        192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251

Db        707 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 766

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311

Db        767 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 826

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371

Db        827 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 886

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTA 431

Db        887 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 946

Qy        432 GGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGC 474

Db        947 CGCGCGCGCGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCGC 989


RESULT 35
ABQ31996/c
ID   ABQ31996 standard; DNA; 1172 BP.
XX
AC   ABQ31996;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 18587.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
```

```
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 1172 BP; 0 A; 583 C; 586 G; 3 T; 0 other;

  Query Match             7.7%;  Score 77.4;  DB 24;  Length 1172;
  Best Local Similarity   47.9%;  Pred. No. 4e-10;
  Matches  222;  Conservative    0;  Mismatches  241;  Indels    0;  Gaps    0;

Qy         12 CGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGC 71
              |||| | |     | | ||  | |   ||| | ||| ||  | | |    || | |||
Db       1144 CGCGCGCGCGCGCGCGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCACGCG 1085

Qy         72 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 131
              ||  ||  || |||  ||  ||  ||    | |   |   ||  ||  || |||  ||
Db       1084 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1025

Qy        132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191
              ||  ||    | |   |   ||  ||  ||    | |   |   ||  ||  || |||
Db       1024 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCACGCGCGCG 965

Qy        192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251
              ||  ||  ||    | |   |   |   ||  ||  ||  |   |   || |||  ||
Db        964 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 905

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
              ||  ||  || |||  ||  ||  ||  || |||  ||  ||  ||  || |||  ||
Db        904 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 845

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
              ||  ||  || |||  ||  ||  ||  || |||  ||  ||  ||  || |||  ||
Db        844 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 785

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTA 431
              ||  ||  || |||  ||  ||  ||  || |||  ||  ||  ||    | |   |
Db        784 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 725

Qy        432 GGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGC 474
                  |   ||  ||  ||    | |   |  |||  ||  |   ||
Db        724 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 682


RESULT 36
ABQ31997
ID   ABQ31997 standard; DNA; 1172 BP.
XX
AC   ABQ31997;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 18588.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 1172 BP; 3 A; 586 C; 583 G; 0 U; 0 other;

  Query Match             7.7%;  Score 77.4;  DB 24;  Length 1172;
  Best Local Similarity   47.9%;  Pred. No. 4e-10;
  Matches  222;  Conservative    0;  Mismatches  241;  Indels    0;  Gaps    0;

Qy         12 CGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGC 71
              |||| | |     | | ||  | |   ||| | ||| ||  | | |    || | |||
Db         29 CGCGCGCGCGCGCGCGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCACGCG 88
```

```
Qy         72 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 131
              ||  ||  || |||  ||  ||  ||    | |  |   ||  ||  || |||  ||
Db         89 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 148

Qy        132 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 191
              ||  ||    | |   |   ||  ||  ||    | |   |   ||  ||  || |||
Db        149 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCACGCGCGCG 208

Qy        192 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 251
              ||  ||  ||    | |   |   ||  ||  ||  |   |   || |||  ||
Db        209 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 268

Qy        252 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 311
              ||  ||  || |||  ||  ||  ||  || |||  ||  ||  ||  || |||  ||
Db        269 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 328

Qy        312 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 371
              ||  ||  || |||  ||  ||  ||  || |||  ||  ||  ||  || |||  ||
Db        329 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 388

Qy        372 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTA 431
              ||  ||  || |||  ||  ||  ||  || |||  ||  ||  ||    | |   |
Db        389 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 448

Qy        432 GGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGC 474
              |   ||  ||  ||    | |   |  |||  ||  |   ||
Db        449 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 491


RESULT 37
AAL06216
ID   AAL06216 standard; DNA; 368 BP.
XX
AC   AAL06216;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 8904.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
```

```
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen
PT   is used in preventing, treating or ameliorating a medical condition  -
XX
PS   Disclosure; SEQ ID NO 8904; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention.
XX
SQ   Sequence 368 BP; 1 A; 99 C; 47 G; 221 T; 0 other;

  Query Match             7.7%;  Score 77.2;  DB 22;  Length 368;
  Best Local Similarity   64.1%;  Pred. No. 3.2e-10;
  Matches  132;  Conservative    0;  Mismatches   73;  Indels    1;  Gaps    1;

Qy        215 AGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 274
              | |  ||||  |||||| |   |  ||||  |  || | ||||||| | || |  | |||
Db         99 ATGCTTGCTTTCTTGCTTGCTTGCTTGCTTTCGTGCTTTCTTGCTTTCCCGTTTTCTTGC 158

Qy        275 TTGCTTGCTCGC-TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 333
              || ||| ||  | |    | ||| |||||| |||  | ||||||||||||  |  || |
Db        159 TTTCTTTCTTTCTTTTGTTTCTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTT 218

Qy        334 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 393
              |||| || |||  |    |  ||  ||  |  ||  ||||||||||| || ||| || ||
Db        219 CTTGTTTTCTCCTTTCTTTCTTTCTTTTGTTTCTTTCCTGCTTGCTTTCTTGCTTGCTTG 278

Qy        394 CTTGCTTGCTCGCTGGCCTGCTTGCT 419
              ||| | ||||  || || | |||| |
Db        279 CTTTCGTGCTTTCTTGCTTTCTTGTT 304


RESULT 38
AAL07359
ID   AAL07359 standard; DNA; 368 BP.
XX
AC   AAL07359;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 10047.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
```

PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition
XX
PS Disclosure; SEQ ID NO 10047; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 368 BP; 1 A; 99 C; 47 G; 221 T; 0 other;

Query Match 7.7%; Score 77.2; DB 22; Length 368;
Best Local Similarity 64.1%; Pred. No. 3.2e-10;
Matches 132; Conservative 0; Mismatches 73; Indels 1; Gaps 1;

```
Qy        215 AGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 274
              | |  ||||  |||||| |   |  ||||  |  || | ||||||| | || |  | |||
Db         99 ATGCTTGCTTTCTTGCTTGCTTGCTTGCTTTCGTGCTTTCTTGCTTTCCCGTTTTCTTGC 158

Qy        275 TTGCTTGCTCGC-TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 333
              || ||| ||  | |    | ||| |||||| |||  | ||||||||||||  |  || |
Db        159 TTTCTTTCTTTCTTTTGTTTCTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTT 218

Qy        334 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 393
              |||| || |||  |    |  ||  ||  |  ||  |||||||||||| || ||| || ||
Db        219 CTTGTTTTCTCCTTTCTTTCTTTCTTTTGTTTCTTTCCTGCTTGCTTTCTTGCTTGCTTG 278

Qy        394 CTTGCTTGCTCGCTGGCCTGCTTGCT 419
              ||| | ||||  || || | |||| |
Db        279 CTTTCGTGCTTTCTTGCTTTCTTGTT 304
```

RESULT 39
ABQ97468/c
ID ABQ97468 standard; cDNA; 392 BP.
XX
AC ABQ97468;
XX
DT 30-OCT-2002 (first entry)
XX
DE Mouse ES cell related cDNA SEQ ID NO 736.
XX
KW Mouse; ES cell; gene trapped sequence; GTS; gene expression;
KW development disorder; cell differentiation disorder; gene; ss.

```
XX
OS   Mus sp.
XX
PN   US2002081668-A1.
XX
PD   27-JUN-2002.
XX
PF   30-NOV-2000; 2000US-0728446.
XX
PR   20-NOV-1998;   98US-109302P.
PR   01-DEC-1999;   99US-168270P.
XX
PA   (FRIE/) FRIEDRICH G.
PA   (ZAMB/) ZAMBROWICZ B.
PA   (SAND/) SANDS A T.
XX
PI   Friedrich G,  Zambrowicz B,  Sands AT;
XX
DR   WPI; 2002-626541/67.
XX
PT   Novel murine polynucleotides that individually identify novel genes
PT   into which a retroviral gene trap vector has been integrated, useful in
PT   genomic analysis and in discovery, development of therapeutic and
PT   diagnostic agents
XX
PS   Claim 2; SEQ ID NO 736; 29pp + Sequence Listing; English.
XX
CC   The invention relates to isolated murine polynucleotides (I) comprising a
CC   contiguous stretch of at least about 60 nucleotides of a sequence
CC   (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences
CC   (GTSs). The novel genes can be used in a process to identify novel
CC   polynucleotide sequences by comparing them to the novel gene sequences.
CC   The novel genes and cells are useful in functional genomic analysis and
CC   in the discovery and development of new therapeutic and diagnostic agents
CC   and methods. (I) is useful for identifying the coding regions of the
CC   murine genome, to isolate cDNAs, genomic clones or full-length
CC   genes/polynucleotides or homologues, heterologues, paralogues or
CC   orthologues that are capable of hybridising to one or more of the GTSs
CC   under stringent conditions. (I) can be incorporated into a phage display
CC   system that can be used to screen for proteins or other ligands, that are
CC   capable of binding an amino acid sequence encoded by an oligonucleotide
CC   or polynucleotide sequence in at least one of the GTS sequences. (I) is
CC   useful in arrays, such as gene chips, to identify and characterise
CC   temporal and tissue specific gene expression, to identify the gene of
CC   interest from many sources and for genetic manipulations such as
CC   antisense inhibition and gene targeting. Decreasing the level of
CC   expression of (I) and/or down regulating the activity of peptides or
CC   proteins encoded by (I) is useful for treating development and cell
CC   differentiation disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at sequdata.uspto.gov/sequence.html?DocID=20020081668.
XX
SQ   Sequence 392 BP; 131 A; 143 C; 63 G; 52 T; 3 other;

  Query Match             7.7%;  Score 77.2;  DB 24;  Length 392;
  Best Local Similarity   59.4%;  Pred. No. 3.2e-10;
  Matches  130;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
              | ||||  | |  |     ||||   |   |||||   |  | ||  | | ||| | | |
Db        223 CCTGCTGTTGGTCCCTGCTGCTGCATTTGCTGCTTTTGCTATTGCTGGATGGCTGGATGG 164

Qy        286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
              ||||  |||| | | ||| | ||||  | ||||| | | |||||   ||| | ||| | |
Db        163 CTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGGTGG 104

Qy        346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 405
              |||  | | ||| | | |||||  | || | | ||| | ||||  | ||| | | | |
Db        103 ATGGCTGGATGGCTGGATGGCTGGTTTNCTGGATGGCTGGATGGCTGGGTTGGTGGATGG 44

Qy        406 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 444
              ||||   ||| | |  | |||  || ||   ||| | ||
Db         43 CTGGATGGCTGGGTTGGTGGATTGTTGGATAGCTGGTTT 5
```

RESULT 40
ABA19579
ID   ABA19579 standard; DNA; 7202 BP.
XX
AC   ABA19579;
XX
DT   23-JAN-2002 (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 11910.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
```
```

PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 11910; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 7202 BP; 1552 A; 1818 C; 1967 G; 1865 T; 0 other;

Query Match 7.7%; Score 76.8; DB 22; Length 7202;
Best Local Similarity 49.0%; Pred. No. 9.7e-10;
Matches 204; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

```
Qy        64 GGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 123
             || | ||  |   |   | | | |||  | | | ||  ||| ||  ||   |  |
Db      1531 GGGAGGGAGGGAAGGAAGATGGGTGGGTGGATGGATGGATGGATAGATGGAAGGGCAGGA 1590

Qy       124 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCT 183
              | |||   |   | || | |||  |  ||   | | | ||   ||  || |||   | |
Db      1591 AGATGGGTGGGGGGGTGGGTGGATGGATGGATGGATCGATGGATGGCTAGAAGGAAGGAT 1650

Qy       184 TGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCT 243
               | |||   |   | |   ||   | | ||  | | | | | | |  |||  | |
Db      1651 GAGAAGATGGATGGGAGGGAGGCAGGGAGGGAAGGAAGATAGGTGGGTGGGTGGGTGGGT 1710

Qy       244 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 303
              | |||   | | | | |   | || | |   | | |   | |||   | | | | | |
Db      1711 AGATGGATGGATGGATAGAAGGAAGGGCCGGAAGATGGGCAGGTGGGTGGGTGGGTGGTT 1770
```

```
Qy        304 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 363
              | |||     | | | | | | |||   | | | | | | | |||   | | | | |||
Db       1771 GGGTGGGTGAATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGCT 1830

Qy        364 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGG 423
              ||||||| ||| ||| ||| ||||||  |||  || |||  |||||  |||
Db       1831 GGCTGGCCGGCTGGCTGGCTGGCTGGCTGTCTGGCTGTCTGGCTGGCTGTAAAGAT 1890

Qy        424 GGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGAT 479
              |||  |   ||  | | | | | | || | | ||   |   |   ||| | | |||
Db       1891 GGAAGGGCAGGAAGGTGGATGGGTGAGAGCAAGGGTGGGTGGGTGGATAGATGGAT 1946


RESULT 41
AAK71750
ID   AAK71750 standard; DNA; 7202 BP.
XX
AC   AAK71750;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:26562.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
```

```
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 26562; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 7202 BP; 1552 A; 1818 C; 1967 G; 1865 T; 0 other;

  Query Match             7.7%;  Score 76.8;  DB 22;  Length 7202;
  Best Local Similarity   49.0%;  Pred. No. 9.7e-10;
  Matches  204;  Conservative    0;  Mismatches  212;  Indels    0;  Gaps    0;

Qy         64 GGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 123
Db       1531 GGGAGGGAGGGAAGGAAGATGGGTGGGTGGATGGATGGATGGATAGATGGAAGGGCAGGA 1590

Qy        124 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCT 183
Db       1591 AGATGGGTGGGGGGGTGGGTGGATGGATGGATGGATCGATGGATGGCTAGAAGGAAGGAT 1650

Qy        184 TGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCT 243
Db       1651 GAGAAGATGGATGGGAGGGAGGCAGGGAGGGAAGGAAGATAGGTGGGTGGGTGGGTGGGT 1710

Qy        244 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 303
Db       1711 AGATGGATGGATGGATAGAAGGAAGGGCCGGAAGATGGGCAGGTGGGTGGGTGGGTGGTT 1770

Qy        304 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 363
Db       1771 GGGTGGGTGAATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGCT 1830

Qy        364 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGG 423
Db       1831 GGCTGGCCGGCTGGCTGGCTGGCTGGCTGGCTGTCTGGCTGTCTGGCTGGCTGTAAAGAT 1890

Qy        424 GGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGAT 479
Db       1891 GGAAGGGCAGGAAGGTGGATGGGTGAGAGCAAGGGTGGGTGGGTGGATAGATGGAT 1946


RESULT 42
ABQ97057/c
ID   ABQ97057 standard; cDNA; 410 BP.
XX
AC   ABQ97057;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Mouse ES cell related cDNA SEQ ID NO 325.
XX
KW   Mouse; ES cell; gene trapped sequence; GTS; gene expression;
KW   development disorder; cell differentiation disorder; gene; ss.
XX
OS   Mus sp.
XX
PN   US2002081668-A1.
XX
PD   27-JUN-2002.
XX
PF   30-NOV-2000; 2000US-0728446.
XX
PR   20-NOV-1998;   98US-109302P.
PR   01-DEC-1999;   99US-168270P.
XX
PA   (FRIE/) FRIEDRICH G.
PA   (ZAMB/) ZAMBROWICZ B.
PA   (SAND/) SANDS A T.
XX
PI   Friedrich G,  Zambrowicz B,  Sands AT;
XX
DR   WPI; 2002-626541/67.
XX
PT   Novel murine polynucleotides that individually identify novel genes
PT   into which a retroviral gene trap vector has been integrated, useful in
PT   genomic analysis and in discovery, development of therapeutic and
PT   diagnostic agents -
XX
PS   Claim 2; SEQ ID NO 325; 29pp + Sequence Listing; English.
XX
CC   The invention relates to isolated murine polynucleotides (I) comprising a
CC   contiguous stretch of at least about 60 nucleotides of a sequence
CC   (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences
CC   (GTSs). The novel genes can be used in a process to identify novel
CC   polynucleotide sequences by comparing them to the novel gene sequences.
CC   The novel genes and cells are useful in functional genomic analysis and
CC   in the discovery and development of new therapeutic and diagnostic agents
CC   and methods. (I) is useful for identifying the coding regions of the
CC   murine genome, to isolate cDNAs, genomic clones or full-length
CC   genes/polynucleotides or homologues, heterologues, paralogues or
CC   orthologues that are capable of hybridising to one or more of the GTSs
CC   under stringent conditions. (I) can be incorporated into a phage display
CC   system that can be used to screen for proteins or other ligands, that are
CC   capable of binding an amino acid sequence encoded by an oligonucleotide
CC   or polynucleotide sequence in at least one of the GTS sequences. (I) is
CC   useful in arrays, such as gene chips, to identify and characterise
CC   temporal and tissue specific gene expression, to identify the gene of
```

```
CC   interest from many sources and for genetic manipulations such as
CC   antisense inhibition and gene targeting. Decreasing the level of
CC   expression of (I) and/or down regulating the activity of peptides or
CC   proteins encoded by (I) is useful for treating development and cell
CC   differentiation disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at sequdata.uspto.gov/sequence.html?DocID=20020081668.
XX
SQ   Sequence 410 BP; 140 A; 151 C; 63 G; 55 T; 1 other;

  Query Match             7.6%;  Score 75.8;  DB 24;  Length 410;
  Best Local Similarity   59.3%;  Pred. No. 7.4e-10;
  Matches  128;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy        226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
              | ||||  | |  |     ||||   |   |||||   |  | ||  | | ||| | | |
Db        241 CCTGCTGTTGGTCCCTGCTGCTGCATTTGCTGCTTTTGCTATTGCTGGATGGCTGGATGG 182

Qy        286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
              ||||  |||| | | ||| | ||||  | ||||| | | |||||   ||| | ||| | |
Db        181 CTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGGTGG 122

Qy        346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 405
               ||||  | | ||| | | |||||  |||| | | ||| | ||||  | ||  | | | |
Db        121 ATGGCTGGATGGCTGGATGGCTGGTTTGCTGGATGGCTGGATGGCTGGGTTCGTGGATGG 62

Qy        406 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTG 441
              ||||   ||| | |  | |||  || ||   ||| |
Db         61 CTGGATGGCTGGGTTGGTGGATTGTTGGATNGCTGG 26


RESULT 43
ABA19578
ID   ABA19578 standard; DNA; 24512 BP.
XX
AC   ABA19578;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 11909.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   20-OCT-2000; 2000US-0242221.
PR   01-NOV-2000; 2000US-0244617.
```

```
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250391.
PR   01-DEC-2000; 2000US-0251160.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system
PT   cancers and metastases -
XX
PS   Disclosure; SEQ ID NO 11909; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 24512 BP; 5908 A; 6031 C; 6107 G; 6466 T; 0 other;

  Query Match             7.5%;  Score 75.4;  DB 22;  Length 24512;
  Best Local Similarity   47.1%;  Pred. No. 3.2e-09;
  Matches  232;  Conservative    0;  Mismatches  261;  Indels    0;  Gaps    0;

Qy       121 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCT 180
            | | | |||   | | | || | || ||| |||   | | | || |  ||    |||
Db     18843 GATGGGTGGGTGGATGGGTGGGAGGGCAGAAGGATGGATAGATGGGTAGATGAGAGGATG 18902

Qy       181 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCT 240
            | | | | | |||   |   | || | |||  | ||||  |  ||   |  ||  |||
Db     18903 GATGGATGGATGGATGGAAAGATGGGTGGATGGGAGGGAGGGAGGGAAGGAAGATGGGTG 18962

Qy       241 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 300
            | | | |||   | | | | | | | |  || | |   | | | | |  ||   | | | |
Db     18963 GGTGGATGGATGGATGGATAGATGGAAGGGCAGGAAGATGGGTGGGGGGGTGGGTGGATG 19022

Qy       301 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 360
            | | | |||   | | | | ||| |  ||   | |     | | | |||   |  ||
Db     19023 GATGGATGGATCGATGGATGGCTAGAAGGAAGGATGAGAAGATGGATGGGAGGGAGGCAG 19082

Qy       361 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 420
            |   |   |   | | | | | | | |||   | | | | | | | |||   | | | |
Db     19083 GGAGGGAAGGAAGATAGGTGGGTGGGTGGGTGGGTAGATGGATGGATGGATGGATGGATA 19142

Qy       421 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 480
               |||  |   ||  | | |   | |  | ||  ||       |   ||||| | |||
Db     19143 GAAGGAAGGGCCGGAAGATGGGCAGGTGGGTGGGTGGGTGAATGGATGGATGGATGGATG 19202

Qy       481 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 540
            |  ||   | |||   | | |  ||   | ||||  | | |  ||   |||||   |||
Db     19203 GATGGATGGATGGATGGATGGATGGATGGATGGGTGGGTGGATGGCTGGCTGGCTGGCTG 19262

Qy       541 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 600
            |  ||   ||||    | | | |||||     ||  | | |  | |     |||  | |
Db     19263 GCTGGCTGGCTGTAAAGATGGAAGGGCAGGAAGGTGGATGGGTGAGAGCAAGGGTGGGTG 19322

Qy       601 GTAGGGCTCCTGG 613
            |  ||     |||
Db     19323 GGTGGATAGATGG 19335


RESULT 44
AAK67571
ID   AAK67571 standard; DNA; 24512 BP.
XX
AC   AAK67571;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:22383.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
```

PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 22383; 3071pp + Sequence Listing; English.

```
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 24512 BP; 5908 A; 6031 C; 6107 G; 6466 T; 0 other;

  Query Match             7.5%;  Score 75.4;  DB 22;  Length 24512;
  Best Local Similarity   47.1%;  Pred. No. 3.2e-09;
  Matches  232;  Conservative    0;  Mismatches  261;  Indels    0;  Gaps    0;

Qy        121 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCT 180
              | | | |||   | | | || | || ||| |||   | | | || |  ||    |||
Db      18843 GATGGGTGGGTGGATGGGTGGGAGGGCAGAAGGATGGATAGATGGGTAGATGAGAGGATG 18902

Qy        181 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCT 240
              | | | | | |||   |   | || | |||  | ||||  |  ||   |  ||  |||
Db      18903 GATGGATGGATGGATGGAAAGATGGGTGGATGGGAGGGAGGGAGGGAAGGAAGATGGGTG 18962

Qy        241 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 300
              | | | |||   | | | | | | | | || | |   | | | | |  ||   | | | |
Db      18963 GGTGGATGGATGGATGGATAGATGGAAGGGCAGGAAGATGGGTGGGGGGGGTGGGTGGATG 19022

Qy        301 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 360
              | | | |||   | | | | ||| |  ||   | |     | | | |||   |   ||
Db      19023 GATGGATGGATCGATGGATGGCTAGAAGGAAGGATGAGAAGATGGATGGGAGGGAGGCAG 19082

Qy        361 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 420
              |   |   |   | | | | | | | |||   | | | | | | | |||   | | | |
Db      19083 GGAGGGAAGGAAGATAGGTGGGTGGGTGGGTGGGTAGATGGATGGATGGATGGATGGATA 19142

Qy        421 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 480
              |||  |   ||  | | |   | |  | ||  ||       |   ||||| | |||
Db      19143 GAAGGAAGGGCCGGAAGATGGGCAGGTGGGTGGGTGGGTGAATGGATGGATGGATGGATG 19202

Qy        481 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 540
              |  ||  | |||   | | |  ||   | ||||  | | |  ||   |||||   |||
Db      19203 GATGGATGGATGGATGGATGGATGGATGGGTGGGTGGATGGCTGGCTGGCTGGCTG 19262

Qy        541 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 600
              |  ||   ||||    | | | |||||     ||  | | |  | |     |||  | |
Db      19263 GCTGGCTGGCTGTAAAGATGGAAGGGCAGGAAGGTGGATGGGTGAGAGCAAGGGTGGGTG 19322

Qy        601 GTAGGGCTCCTGG 613
              |  ||     |||
Db      19323 GGTGGATAGATGG 19335


RESULT 45
AAK71749
ID   AAK71749 standard; DNA; 24512 BP.
XX
AC   AAK71749;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:26561.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
```

```
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 26561; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 24512 BP; 5908 A; 6031 C; 6107 G; 6466 T; 0 other;

  Query Match             7.5%;  Score 75.4;  DB 22;  Length 24512;
  Best Local Similarity   47.1%;  Pred. No. 3.2e-09;
  Matches  232;  Conservative    0;  Mismatches  261;  Indels    0;  Gaps    0;

Qy        121 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCT 180
              | | | |||   | | | || | || ||| |||   | | | || | |  ||    |||
Db      18843 GATGGGTGGGTGGATGGGTGGGAGGGCAGAAGGATGGATAGATGGGTAGATGAGAGGATG 18902

Qy        181 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCT 240
              | | | | | |||   |   | || | |||  | ||||  |  ||   |  ||  |||
Db      18903 GATGGATGGATGGATGGAAAGATGGGTGGATGGGAGGGAGGGAGGGAAGGAAGATGGGTG 18962

Qy        241 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 300
              | | | |||   | | | | | | |  || | |   | | | | |  ||   | | | |
Db      18963 GGTGGATGGATGGATGGATAGATGGAAGGGCAGGAAGATGGGTGGGGGGGTGGGTGGATG 19022

Qy        301 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 360
              | | | |||   | | | | ||| |  ||   | |     | | | |||   |   ||
Db      19023 GATGGATGGATCGATGGATGGCTAGAAGGAAGGATGAGAAGATGGATGGGAGGGAGGCAG 19082

Qy        361 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 420
              |   |   |   | | | | | | | | |||   | | | | | | | |||   | | | |
Db      19083 GGAGGGAAGGAAGATAGGTGGGTGGGTGGGTGGGTAGATGGATGGATGGATGGATGGATA 19142

Qy        421 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 480
              |||  |   ||  | | |   | |  | ||  ||       |   ||||| | |||
Db      19143 GAAGGAAGGGCCGGAAGATGGGCAGGTGGGTGGGTGGGTGAATGGATGGATGGATGGATG 19202

Qy        481 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 540
              |  ||   | |||  | | |  ||   | ||||  | | |  ||   |||||   |||
Db      19203 GATGGATGGATGGATGGATGGATGGATGGGTGGGTGGATGGCTGGCTGGCTGGCTG 19262

Qy        541 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 600
              |  ||   ||||    | | | |||||     ||  | | |  | |     |||  | |
Db      19263 GCTGGCTGGCTGTAAAGATGGAAGGGCAGGAAGGTGGATGGGTGAGAGCAAGGGTGGGTG 19322

Qy        601 GTAGGGCTCCTGG 613
              |  ||     |||
Db      19323 GGTGGATAGATGG 19335
```

```
RESULT 46
ABQ96982/c
ID   ABQ96982 standard; cDNA; 427 BP.
XX
AC   ABQ96982;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Mouse ES cell related cDNA SEQ ID NO 250.
XX
KW   Mouse; ES cell; gene trapped sequence; GTS; gene expression;
KW   development disorder; cell differentiation disorder; gene; ss.
XX
OS   Mus sp.
XX
PN   US2002081668-A1.
XX
PD   27-JUN-2002.
XX
PF   30-NOV-2000; 2000US-0728446.
XX
PR   20-NOV-1998;   98US-109302P.
PR   01-DEC-1999;   99US-168270P.
XX
PA   (FRIE/) FRIEDRICH G.
PA   (ZAMB/) ZAMBROWICZ B.
PA   (SAND/) SANDS A T.
XX
PI   Friedrich G,  Zambrowicz B,  Sands AT;
XX
DR   WPI; 2002-626541/67.
XX
PT   Novel murine polynucleotides that individually identify novel genes
PT   into which a retroviral gene trap vector has been integrated, useful in
PT   genomic analysis and in discovery, development of therapeutic and
PT   diagnostic agents -
XX
PS   Claim 2; SEQ ID NO 250; 29pp + Sequence Listing; English.
XX
CC   The invention relates to isolated murine polynucleotides (I) comprising a
CC   contiguous stretch of at least about 60 nucleotides of a sequence
CC   (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences
CC   (GTSs). The novel genes can be used in a process to identify novel
CC   polynucleotide sequences by comparing them to the novel gene sequences.
CC   The novel genes and cells are useful in functional genomic analysis and
CC   in the discovery and development of new therapeutic and diagnostic agents
CC   and methods. (I) is useful for identifying the coding regions of the
CC   murine genome, to isolate cDNAs, genomic clones or full-length
CC   genes/polynucleotides or homologues, heterologues, paralogues or
CC   orthologues that are capable of hybridising to one or more of the GTSs
CC   under stringent conditions. (I) can be incorporated into a phage display
CC   system that can be used to screen for proteins or other ligands, that are
CC   capable of binding an amino acid sequence encoded by an oligonucleotide
CC   or polynucleotide sequence in at least one of the GTS sequences. (I) is
CC   useful in arrays, such as gene chips, to identify and characterise
CC   temporal and tissue specific gene expression, to identify the gene of
CC   interest from many sources and for genetic manipulations such as
CC   antisense inhibition and gene targeting. Decreasing the level of
CC   expression of (I) and/or down regulating the activity of peptides or
CC   proteins encoded by (I) is useful for treating development and cell
CC   differentiation disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at sequdata.uspto.gov/sequence.html?DocID=20020081668.
XX
SQ   Sequence 427 BP; 148 A; 155 C; 66 G; 54 T; 4 other;

  Query Match             7.4%;  Score 74.4;  DB 24;  Length 427;
  Best Local Similarity   58.3%;  Pred. No. 1.7e-09;
  Matches  126;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy       226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
            | ||||  | |  |    ||||   |    ||||   |  | ||  | | ||| | |
Db       258 CCTGCTGTTGGTCCCTGCTGCTGCATTTGCNGCTTTTGCTATTGCTGGATGGCTGGATGN 199

Qy       286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
             ||||  |||| | | ||| | ||||  | |||||  | |  |||||   ||| | ||| |
Db       198 CTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGGTGN 139

Qy       346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 405
              ||||  | | ||| | | |||||  ||||   | ||| | ||||  | ||| | | | |
Db       138 ATGGCTGGATGGCTGGATGGCTGGTTTGCTGNATGGCTGGATGGCTGGGTTGGTGGATGG 79

Qy       406 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTG 441
             ||||   ||| | |  | |||  || ||  |||| |
Db        78 CTGGATGGCTGGGTTGGTGGATTGTTGGATTGCTGG 43


RESULT 47
ABQ96978/c
ID   ABQ96978 standard; cDNA; 379 BP.
XX
AC   ABQ96978;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Mouse ES cell related cDNA SEQ ID NO 246.
XX
KW   Mouse; ES cell; gene trapped sequence; GTS; gene expression;
KW   development disorder; cell differentiation disorder; gene; ss.
XX
OS   Mus sp.
XX
PN   US2002081668-A1.
XX
PD   27-JUN-2002.
XX
PF   30-NOV-2000; 2000US-0728446.
XX
PR   20-NOV-1998;   98US-109302P.
PR   01-DEC-1999;   99US-168270P.
XX
PA   (FRIE/) FRIEDRICH G.
PA   (ZAMB/) ZAMBROWICZ B.
PA   (SAND/) SANDS A T.
XX
PI   Friedrich G,  Zambrowicz B,  Sands AT;
XX
DR   WPI; 2002-626541/67.
XX
PT   Novel murine polynucleotides that individually identify novel genes
PT   into which a retroviral gene trap vector has been integrated, useful in
PT   genomic analysis and in discovery, development of therapeutic and
PT   diagnostic agents -
XX
PS   Claim 2; SEQ ID NO 246; 29pp + Sequence Listing; English.
XX
CC   The invention relates to isolated murine polynucleotides (I) comprising a
CC   contiguous stretch of at least about 60 nucleotides of a sequence
CC   (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences
CC   (GTSs). The novel genes can be used in a process to identify novel
CC   polynucleotide sequences by comparing them to the novel gene sequences.
CC   The novel genes and cells are useful in functional genomic analysis and
CC   in the discovery and development of new therapeutic and diagnostic agents
CC   and methods. (I) is useful for identifying the coding regions of the
CC   murine genome, to isolate cDNAs, genomic clones or full-length
CC   genes/polynucleotides or homologues, heterologues, paralogues or
CC   orthologues that are capable of hybridising to one or more of the GTSs
CC   under stringent conditions. (I) can be incorporated into a phage display
CC   system that can be used to screen for proteins or other ligands, that are
CC   capable of binding an amino acid sequence encoded by an oligonucleotide
CC   or polynucleotide sequence in at least one of the GTS sequences. (I) is
CC   useful in arrays, such as gene chips, to identify and characterise
CC   temporal and tissue specific gene expression, to identify the gene of
CC   interest from many sources and for genetic manipulations such as
```

```
CC   antisense inhibition and gene targeting. Decreasing the level of
CC   expression of (I) and/or down regulating the activity of peptides or
CC   proteins encoded by (I) is useful for treating development and cell
CC   differentiation disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at sequdata.uspto.gov/sequence.html?DocID=20020081668.
XX
SQ   Sequence 379 BP; 120 A; 143 C; 61 G; 50 T; 5 other;

  Query Match             7.3%;  Score 72.8;  DB 24;  Length 379;
  Best Local Similarity   57.9%;  Pred. No. 4.2e-09;
  Matches  125;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy        226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
Db        239 CCTGCTGTTGGTCCCTGCTGCTGCATTTGCTGCTTTTGCTATTGCTGGATGGCTGGATGG 180

Qy        286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
Db        179 CTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGNTGGTGN 120

Qy        346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 405
Db        119 ATGNCTGNATGGCTGGATGGCTGGTTTGCTGGATGGCTGGATGGCTGGGTTGGTGGATGG 60

Qy        406 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTG 441
Db         59 CTGGATGGCTGGGTTGGTGGATTGTTAGATTGCTGG 24


RESULT 48
AAV88624/c
ID   AAV88624 standard; cDNA; 368 BP.
XX
AC   AAV88624;
XX
DT   12-FEB-1999  (first entry)
XX
DE   EST clone GX576.
XX
KW   Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW   tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW   chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW   receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS   Homo sapiens.
XX
PN   WO9845437-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06956.
XX
PR   10-APR-1997;   97US-0837312.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070078/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries
XX
PS   Claim 1; Page 458; 641pp; English.
XX
CC   The present sequence represents an expressed sequence tag (EST), and is
CC   a polynucleotide of the invention. The polynucleotides of the invention
CC   are all secreted EST sequences isolated from a variety of human tissue
CC   sources. The EST sequences and proteins encoded by them are predicted to
CC   have useful biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals, although no supporting data is given. Suggested activities
CC   include nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The EST sequences are also stated to be useful for gene
CC   therapy.
XX
SQ   Sequence 368 BP; 216 A; 51 C; 89 G; 9 T; 3 other;

  Query Match             7.3%;  Score 72.6;  DB 20;  Length 368;
  Best Local Similarity   65.2%;  Pred. No. 4.7e-09;
  Matches  137;  Conservative    0;  Mismatches   67;  Indels    6;  Gaps    2;

Qy        220 TGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 279
Db        227 TTCTTGCTTGCTTTCTTGCTTGCTNGCTTGCTTTCGTGCTTTCTTGCTTTCCTGTTTNCT 168

Qy        280 TGCTCGCTGGCCTGCTTG-----CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 334
Db        167 TTCTTTCTTTCTTTCTTTTGTTTCTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTCGTGC 108

Qy        335 TTGCTTGCTCGCTGGCCTGCTTGCTTGC-TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 393
Db        107 TTTCTTGTTTTCTCGCNTTCTTTCTTTCTTTTGTTTCTTTCCTGCTTGCTTTCTTGCTTG 48

Qy        394 CTTGCTTGCTCGCTGGCCTGCTTGCTGAGG 423
Db         47 ATTGCTTTCGTGCTTTCTTGCTTTCTTAGG 18


RESULT 49
AAK69442
ID   AAK69442 standard; DNA; 16084 BP.
XX
AC   AAK69442;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24254.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
```

PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 24254; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic

```
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 16084 BP; 2716 A; 4761 C; 5154 G; 3453 T; 0 other;

  Query Match             7.2%;  Score 71.8;  DB 22;  Length 16084;
  Best Local Similarity   49.9%;  Pred. No. 2.3e-08;
  Matches  211;  Conservative    0;  Mismatches  207;  Indels    5;  Gaps    1;

Qy        260 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 319
              ||  ||| ||   |  ||       |  |||||    ||| ||| |  |||    |  |
Db       8290 TGAGCGCAGGAGCGTCTGGGCCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCTTGTCC 8349

Qy        320 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 379
              |    |  |||||    ||| ||| |  |||    |  | |    |  |||||    |||
Db       8350 TCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCTTGTCCTCACGGTGGGCCTATCAGCT 8409

Qy        380 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 439
              ||| |  |||      |||||  | | ||   |     | ||  | | | ||||| |||
Db       8410 GGCTGGAGGGC-----GGCTTGTCCTCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCT 8464

Qy        440 TGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCT 499
              ||    |   | || |    |  ||| |||  ||   | | ||      | |||||   |
Db       8465 TGTCCTCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCTTGTCCTCAGGGTGGGCCTAT 8524

Qy        500 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 559
              |||||  |||| ||| ||      | |||||   |     |||||  |||| |||
Db       8525 CAGCTGGCTGGAGGGCGGCTTGTCCTCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCT 8584

Qy        560 AGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCT 619
              ||       |||||   |  ||||| |||| ||| ||       |||||   |
Db       8585 TGTCCTCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCTTGTCCTCAGGGTGGGCCTAT 8644

Qy        620 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 679
              |||||  |||| ||| ||      | |||||   |     |||||  |||| |||
Db       8645 CAGCTGGCTGGAGGGCGGCTTGTCCTCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCT 8704

Qy        680 AGT 682
              ||
Db       8705 TGT 8707


RESULT 50
AAK82881/c
ID   AAK82881 standard; DNA; 16084 BP.
XX
AC   AAK82881;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:37693.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
```

PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 37693; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 16084 BP; 3453 A; 5154 C; 4761 G; 2716 T; 0 other;

Query Match 7.2%; Score 71.8; DB 22; Length 16084;
Best Local Similarity 49.9%; Pred. No. 2.3e-08;
Matches 211; Conservative 0; Mismatches 207; Indels 5; Gaps 1;

```
Qy       260 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 319
             ||  ||| ||   |  ||       | |||||    ||| ||| |  |||    |  |
Db      7795 TGAGCGCAGGAGCGTCTGGGCCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCTTGTCC 7736

Qy       320 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 379
             |    | |||||    ||| ||| |  |||    |  | |    |  |||||    |||
Db      7735 TCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCTTGTCCTCACGGTGGGCCTATCAGCT 7676

Qy       380 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 439
              ||| |  |||       |||||  | | ||   |     | ||  | | | ||||| |||
Db      7675 GGCTGGAGGGC-----GGCTTGTCCTCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCT 7621

Qy       440 TGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCT 499
             ||    |   | || |    |  ||| |||  ||   | | ||      | |||||   |
Db      7620 TGTCCTCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCTTGTCCTCAGGGTGGGCCTAT 7561

Qy       500 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 559
                 |||||  |||| ||| ||      | |||||   |     |||||  |||| |||
Db      7560 CAGCTGGCTGGAGGGCGGCTTGTCCTCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCT 7501

Qy       560 AGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCT 619
              ||         |||||   |     |||||  |||| ||| ||        |||||   |
Db      7500 TGTCCTCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCTTGTCCTCAGGGTGGGCCTAT 7441

Qy       620 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 679
                 |||||  |||| ||| ||      | |||||   |     |||||  |||| |||
Db      7440 CAGCTGGCTGGAGGGCGGCTTGTCCTCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCT 7381

Qy       680 AGT 682
              ||
Db      7380 TGT 7378
```

RESULT 51
AAL40272/c
ID AAL40272 standard; DNA; 157 BP.
XX
AC AAL40272;
XX
DT 13-SEP-2002 (first entry)
XX
DE BASB204 related DNA sequence SEQ ID No 33.
XX
KW Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW auditive nerve damage; delayed speech learning; gene; ds.
XX
OS Haemophilus influenzae.
XX
PN WO200234768-A2.
XX
PD 02-MAY-2002.

```
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   P-PSDB; AAO21919.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear  -
XX
PS   Example 2; Fig 3; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 157 BP; 66 A; 33 C; 28 G; 30 T; 0 other;

  Query Match             7.1%;  Score 70.8;  DB 24;  Length 157;
  Best Local Similarity   86.7%;  Pred. No. 1.1e-08;
  Matches   78;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        251 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 310
              ||||||||||||| ||| || |||||||||||| ||| || |||||||||||| ||| ||
Db        100 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 41

Qy        311 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 340
              ||||||||||||| ||| || |||||||||
Db         40 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 11


RESULT 52
AAN50853
ID   AAN50853 standard; DNA; 600 BP.
XX
AC   AAN50853;
XX
DT   15-JAN-1992  (first entry)
XX
DE   Sequence encoding tetra-gastrin polymer.
XX
KW   E.coli; gastrin; ds.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   repeat_unit     1..12
FT                   /*tag= a
FT                   /number= 1
FT                   /note= "Repeated from 2-50 times"
XX
PN   JP60188081-A.
XX
PD   25-SEP-1985.
XX
PF   09-MAR-1984;   84JP-0046180.
XX
PR   09-MAR-1984;   84JP-0046180.
XX
PA   (SUMO ) SUMITOMO CHEM IND KK.
XX
DR   WPI; 1985-279157/45.
DR   P-PSDB; AAP50588.
XX
PT   Tetra-gastrin polymer gene - comprises many nucleotide sequence
PT   units of tetra-gastrin aminoacid sequence.
XX
PS   Claim 2; Page 491; 7pp; Japanese.
XX
CC   The repeat sequence encodes the tetra-gastrin polymer, and is useful
CC   in the large scale production of that protein, pref. through an E.
CC   coli host transformed with a pBR322 construct.
XX
SQ   Sequence 600 BP; 100 A; 100 C; 200 G; 200 T; 0 other;

  Query Match             7.1%;  Score 70.6;  DB 6;  Length 600;
  Best Local Similarity   44.9%;  Pred. No. 1.8e-08;
  Matches  268;  Conservative    0;  Mismatches  329;  Indels    0;  Gaps    0;

Qy        127 TGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGC 186
              |||   |  | |||   ||||    ||   |  | |||   ||||    ||   |  | |
Db          1 TGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTC 60

Qy        187 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGC 246
              | | ||| || || | ||         |  || || |||| | |      ||  |   |
Db         61 TGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTC 120

Qy        247 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 306
              |||   | | || | | |   | | | |  | || | ||| || || | | |   |
Db        121 TGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTC 180

Qy        307 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 366
              |||   |  | || | | |   | | | |  | || | ||| || || | | |   |
Db        181 TGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTC 240

Qy        367 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 426
              |||   |  | || | | |   | | | |  | || | ||| || || | | ||
Db        241 TGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTC 300

Qy        427 CAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGG 486
              |  || || || | | |     ||| ||    |  |   ||   || ||| |
Db        301 TGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTC 360

Qy        487 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 546
              | ||| || || |  || || ||||  |      ||   |       | | |
Db        361 TGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTC 420

Qy        547 CTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 606
              | ||| || || |  || || ||||  |      ||   |       | | |
Db        421 TGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTC 480

Qy        607 CTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 666
              |||| || || |  || || ||||  |      ||   |       | | |
Db        481 TGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTC 540

Qy        667 CTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGC 723
              | ||| || || |  || || |  | |||  | || | | |  | || |  || |
Db        541 TGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGAC 597
```

```
RESULT 53
AAS76504
ID   AAS76504 standard; cDNA; 3249 BP.
XX
AC   AAS76504;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #12308.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG12317.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 1; SEQ ID No 12308; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3249 BP; 631 A; 1058 C; 972 G; 588 T; 0 other;

  Query Match             7.0%;  Score 69.6;  DB 23;  Length 3249;
  Best Local Similarity   50.0%;  Pred. No. 5.3e-08;
  Matches  174;  Conservative    0;  Mismatches  174;  Indels    0;  Gaps    0;

Qy        146 ACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGA 205
              |||| | | ||| | ||   |    |||   ||||| | ||  | |  || || ||
Db       2858 ACAGCATGTCTGATGGCGTGGTCCCCAGCCTGCCTGATGGCGTGGTCCCCAGCCTGTCTG 2917

Qy        206 GGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCG 265
                ||   |     | ||  |  || | |   ||   |  |   | ||| | || || ||
Db       2918 ATGGCATGGTCCCCGGCCTGTCTGATGGCGTGGTCCCCAGTCTGTCTGATGGCGTGGTCC 2977

Qy        266 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 325
              | |||||  || | ||  | |  || || ||  || | ||  |    |  |  ||  |
Db       2978 CCGGCCTGTCTGATGGCGTGGTCCCCGGCCTGTCTGATGGCATGGTCCCCGGTCTGTCTG 3037

Qy        326 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 385
               |||  || |  |  ||  |   |   |  || |  |  ||  | ||| | ||||| ||
Db       3038 ATGGTGTGGTCCCCAGCCTGTCTGATGGTGTGGTCCCCGGCCTGTCTGATGGCTTGGTCC 3097

Qy        386 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTG 445
              |||| ||  ||| | |   | |  || || ||      ||   |     | || ||  ||
Db       3098 CTGGTCTATTTGATGGTGTGGTCCCCAGCCTGTCTGAAGGTGTGGTCCCCAGCCTGTCTG 3157

Qy        446 CTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 493
                |   | |       ||||| |||||||||||| |    || ||  | |
Db       3158 ATGGCGTGGTCCCCGGCCTGTCTGATGGCTTGGTCCCTGGTCTATTTG 3205


RESULT 54
AAK85846
ID   AAK85846 standard; DNA; 7614 BP.
XX
AC   AAK85846;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40658.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
```

PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis - .
XX
PS Disclosure; SEQ ID NO 40658; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 7614 BP; 2213 A; 1415 C; 1685 G; 2301 T; 0 other;

Query Match 6.9%; Score 69.4; DB 22; Length 7614;
Best Local Similarity 46.8%; Pred. No. 7.6e-08;
Matches 318; Conservative 0; Mismatches 356; Indels 5; Gaps 3;

Qy 25 GGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTC 84
Db 3029 GGGTGCAGGTGATGCTGATTGGGTGCAGGTGATGCTGGCTGAATGGGTGCAGGTGATGCT 3088

Qy 85 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 144
Db 3089 GATTGGGTGCAGGTGATGCTGATTGACTGCGTGCAGG---TGATGCTAGCTGATTGGGTG 3145

```
Qy        145 GACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 204
Db       3146 CAGGTGATGCTGGCTGACTGGGTGCAGGTGATGTTGGCTGATTGGGTGCAGGTGATGCTG 3205

Qy        205 AGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTC 264
Db       3206 ATTGGGTGCAGGTGATGCTGGCTGATTGGGTGCAGGTGATGCTGGCTGATTGGGTGCAGG 3265

Qy        265 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 324
Db       3266 TGTTAGCTGATTGGGTGCAGGTGATGCTGATTGGGTGCAGGTGATGCTGGCTGACTGGAT 3325

Qy        325 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 384
Db       3326 GCAAGTGATGCTGGCCGATTGGGTGCAGGTGATGCTGGATGACTGGGTGCAGGTGATGCT 3385

Qy        385 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 444
Db       3386 AGATGACTGGGTGCAGGCGATGTTGGCTGATTGGGTGCAGGTGATGTTGGCTGATTGGGT 3445

Qy        445 GCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAG 504
Db       3446 GCAGGTGATGCTGGCTGTGTGCAGGTGATGCTGATTGGGTGCAGGTGGTGCTGGTTGACT 3505

Qy        505 GGCTGCTGG-GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 563
Db       3506 GGGTGCAGGTGCTGCTGATTGGGTGCAGGTGATGCTGATTGGATGCAGGTGATGCTGATT 3565

Qy        564 GGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTA 623
Db       3566 GGGTGCAGGTGATGCTGGT-TGATTGGGTGCAGGTGATGCTGGCTGACTGGATGCAGGTG 3624

Qy        624 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 683
Db       3625 ATGCTGATTGGGTGCAGGTGATGCTGATTGGATGCAGGTGATGCTGATTGGGTGCAGGTG 3684

Qy        684 GGGCTGCCTGCTGGCTTGC 702
Db       3685 ATGCTGGCTGATTGGGTGC 3703
```

```
RESULT 55
AAS46592
ID   AAS46592 standard; DNA; 6286 BP.
XX
AC   AAS46592;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #314.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW   cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP02955.
XX
PR   15-MAR-2000; 2000DE-1013847.
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumour suppressor
PT   genes and oncogenes, useful in designing primers and probes for
PT   analysing diseases associated with cytosine methylation state e.g.
PT   cancer -
XX
PS   Claim 1; SEQ ID No 314; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and
CC   oncogenes having a sequence taken from 536 (actually 533 since
CC   numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC   (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC   peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC   form part of a set of probes for detecting the cytosine methylation state
CC   and/or single nucleotide polymorphisms and also to be used in an
CC   array for analysing diseases associated with CpG dinucleotides e.g.
CC   cancers and tumours. The probes can also be used in a method for
CC   ascertaining genetic and/or epigenetic parameters for the diagnosis
CC   and/or therapy of existing diseases or the predisposition to specific
CC   diseases, by analysing cytosine methylations. The parameters may be
CC   compared to another set of genetic and/or epigenetic parameters, the
CC   differences serving as basis for diagnosis and/or prognosis events which
CC   are disadvantageous to patients. The present sequence is one of the
CC   533 genomic sequences derived from tumour suppressor genes and
CC   oncogenes. Sequences with even numbered Seq ID numbers are the
CC   complementary sequence of the corresponding odd numbered sequence (e.g.
CC   ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC   is missing).
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 6286 BP; 1526 A; 124 C; 1714 G; 2922 T; 0 other;

  Query Match             6.9%;  Score 69.2;  DB 22;  Length 6286;
  Best Local Similarity   45.7%;  Pred. No. 8.1e-08;
  Matches  317;  Conservative    0;  Mismatches  373;  Indels    4;  Gaps    2;

Qy         76 TGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCT 135
Db       1760 TGGTGGATGGATGGATGGATGGATGGATGTATAGGTGGATGGATGGATGGATGGATGGAT 1819

Qy        136 TGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 195
Db       1820 GGATGGATGGATGGATGGATGTATAAGTGGATGGATGGATGGATGGATGGATGGATGGAT 1879

Qy        196 TGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCT 255
Db       1880 GGATGGATGGATGTATAGGTGGATGGATGGATGGATGGATGGATGGATGGATGGATGGAT 1939

Qy        256 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 315
Db       1940 GGATGTATAAGTGGATGGATGGATGGATGGATGGATGGATGTATAAGTGGATGGATGGAT 1999

Qy        316 TGCTTGCTCGCTGGCC---TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 372
Db       2000 GGATGGATGGATGGGTGAGTAGTAGGTGAGTGGGTAAATGGGTGGGTGGATGGGTGGATG 2059

Qy        373 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAG 432
Db       2060 GATGGATGTATAGATGGATGGATGGATGGATGTATAGGTGGATGGATGGATGGATGGATG 2119

Qy        433 GGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTG 492
Db       2120 GATGGATGGATGTATAAGTGGATGGATGGATGGATGGATGGATGGAT-GGATGGATGGAT 2178

Qy        493 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 552
```

```
Db       2179 GGATGGATGTATAGATGGATGGATGGATGGATGGATGGATGGATGTATAAATGGAT 2238

Qy        553 GGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTG 612

Db       2239 GGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGTATAGGTGGATGGAT 2298

Qy        613 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 672

Db       2299 GGATGGATGGATGGATGGATGGATGGATGTATAAGTGGATGGATGGATGTATAAGT 2358

Qy        673 GGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCT 732

Db       2359 GGATGGATGGATGGATGGATGGATGGATGGATGGATGGGTGAGTAGTAGGTGAGTG 2418

Qy        733 GCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGA 766

Db       2419 GGTAAATGGGTGGGTGGATGGGTGGATGGATGGA 2452


RESULT 56
AAS70903/c
ID   AAS70903 standard; cDNA; 3588 BP.
XX
AC   AAS70903;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #6707.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG06716.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 6707; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3588 BP; 930 A; 1100 C; 709 G; 849 T; 0 other;

  Query Match             6.7%;  Score 67.2;  DB 23;  Length 3588;
  Best Local Similarity   44.6%;  Pred. No. 2.2e-07;
  Matches  311;  Conservative    0;  Mismatches  383;  Indels    4;  Gaps    1;

Qy         44 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 103

Db       1652 TGGATGCAAGTGATGCTGGCCGATTGGGTGCAGGTGATGCTGGATGACTGGGTGCAGGTG 1593

Qy        104 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 163

Db       1592 ATGCTAGATGACTGGGTGCAGGCGATGTTGGCTGATTGGGTGCAGGTGATGTTGGCTGAT 1533

Qy        164 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 223

Db       1532 TGGGTGCAGGTGATGCTGGCTGTGTGCAGGTGATGCTGATTGGGTGCAGGTGGTGCTGGT 1473

Qy        224 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 283

Db       1472 TGACTGGGTGCAGGTGCTGCTGATTGGGTGCAGGTGATGCTGATTGGATGCAGGTGATGC 1413

Qy        284 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 343

Db       1412 TGATTGGGTGCAGGTGATGCTGGTTGATTGGGTGCAGGTGATGCTGGCTGACTGGATGCA 1353

Qy        344 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403

Db       1352 GGTGATGCTGATTGGGTGCAGGTGATGCTGATTGGATGCAGGTGATGCTGATTGGGTGCA 1293

Qy        404 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCC 463

Db       1292 GGTGATGCTGGCTGATTGGGTGCGGGTGATGCTGGCTGATTG----GGTGCAGATGATGC 1237

Qy        464 TGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 523

Db       1236 TGGCTGATTGGGTGCAAGTGATGCTGATTGATTGGGTGCAGGTGATGCTGGCTGATTGGG 1177

Qy        524 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTA 583

Db       1176 TGCAGATGATGCTGGCTGATTGGGTGTGGGTGATGCTGGCTGATTGGGTGCAGGTGATGC 1117

Qy        584 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 643

Db       1116 TGGCTGACTGGGTGCAGGTGATGCTGATTGCTTGGGTGCAGGTGATGCTGGCTGATTGGA 1057

Qy        644 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCT 703

Db       1056 TGCAGGTGATGCTGATTGTGTGCAGGTGATGCTGGATGACTGGGTGCAGGTGATGCTGCA 997

Qy        704 TGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGG 741

Db        996 TGATTGAGTGCAGGTGATGCTGGCTGATTGGGTGCAGG 959


RESULT 57
AAS92599/c
ID   AAS92599 standard; cDNA; 3588 BP.
XX
AC   AAS92599;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #28403.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
```

```
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG28412.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 28403; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3588 BP; 930 A; 1100 C; 709 G; 849 T; 0 other;

  Query Match             6.7%;  Score 67.2;  DB 23;  Length 3588;
  Best Local Similarity   44.6%;  Pred. No. 2.2e-07;
  Matches  311;  Conservative    0;  Mismatches  383;  Indels    4;  Gaps    1;

Qy         44 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 103
              ||  |||  ||||  |     |   || |||  |       |   | |||  ||| |  |
Db       1652 TGGATGCAAGTGATGCTGGCCGATTGGGTGCAGGTGATGCTGGATGACTGGGTGCAGGTG 1593

Qy        104 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 163
                |   | | |||  |||  ||     ||  || |  || |   ||     ||  || |
Db       1592 ATGCTAGATGACTGGGTGCAGGCGATGTTGGCTGATTGGGTGCAGGTGATGTTGGCTGAT 1533

Qy        164 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 223
               || |   ||    |||  ||   || ||      || |  || |   || | |||| |
Db       1532 TGGGTGCAGGTGATGCTGGCTGTGTGCAGGTGATGCTGATTGGGTGCAGGTGGTGCTGGT 1473

Qy        224 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 283
              |  ||   | |||       | | |  |||  |       | | |  |||  |
Db       1472 TGACTGGGTGCAGGTGCTGCTGATTGGGTGCAGGTGATGCTGATTGGATGCAGGTGATGC 1413

Qy        284 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 343
              | | |  |||  |       | | |  ||  |||  |       | |||  ||  |||
Db       1412 TGATTGGGTGCAGGTGATGCTGGTTGATTGGGTGCAGGTGATGCTGGCTGACTGGATGCA 1353

Qy        344 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 403
              |     ||| |||  |||  |     ||| |||  |||  |     ||| |||  |||
Db       1352 GGTGATGCTGATTGGGTGCAGGTGATGCTGATTGGATGCAGGTGATGCTGATTGGGTGCA 1293

Qy        404 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCC 463
              |     |||  || |  || |   ||   ||||  || |||    || |  | |    |
Db       1292 GGTGATGCTGGCTGATTGGGTGCGGGTGATGCTGGCTGATTG----GGTGCAGATGATGC 1237

Qy        464 TGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 523
              || ||||| |  ||  |||   |||| | |  ||   | |||       ||||| | |
Db       1236 TGGCTGATTGGGTGCAAGTGATGCTGATTGATTGGGTGCAGGTGATGCTGGCTGATTGGG 1177

Qy        524 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTA 583
              |        ||||   |   || ||  ||       |   | |  |   |   | |
Db       1176 TGCAGATGATGCTGGCTGATTGGGTGTGGGTGATGCTGGCTGATTGGGTGCAGGTGATGC 1117

Qy        584 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 643
              |||||   || |||  ||   |||  | |  ||   | |||       ||||| | | |
Db       1116 TGGCTGACTGGGTGCAGGTGATGCTGATTGCTTGGGTGCAGGTGATGCTGGCTGATTGGA 1057

Qy        644 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCT 703
              |   |    |||| | ||  |   |    ||||   |   || |||  |       |||
Db       1056 TGCAGGTGATGCTGATTGTGTGCAGGTGATGCTGGATGACTGGGTGCAGGTGATGCTGCA 997

Qy        704 TGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGG 741
              || |||  |||  |||  || |  ||  ||      |||
Db        996 TGATTGAGTGCAGGTGATGCTGGCTGATTGGGTGCAGG 959


RESULT 58
ABQ97197
ID   ABQ97197 standard; cDNA; 397 BP.
XX
AC   ABQ97197;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Mouse ES cell related cDNA SEQ ID NO 465.
XX
KW   Mouse; ES cell; gene trapped sequence; GTS; gene expression;
KW   development disorder; cell differentiation disorder; gene; ss.
XX
OS   Mus sp.
XX
PN   US2002081668-A1.
XX
PD   27-JUN-2002.
XX
PF   30-NOV-2000; 2000US-0728446.
XX
PR   20-NOV-1998;   98US-109302P.
PR   01-DEC-1999;   99US-168270P.
XX
PA   (FRIE/) FRIEDRICH G.
PA   (ZAMB/) ZAMBROWICZ B.
PA   (SAND/) SANDS A T.
XX
PI   Friedrich G,  Zambrowicz B,  Sands AT;
XX
DR   WPI; 2002-626541/67.
XX
PT   Novel murine polynucleotides that individually identify novel genes
PT   into which a retroviral gene trap vector has been integrated, useful in
PT   genomic analysis and in discovery, development of therapeutic and
PT   diagnostic agents  -
XX
PS   Claim 2; SEQ ID NO 465; 29pp + Sequence Listing; English.
XX
CC   The invention relates to isolated murine polynucleotides (I) comprising a
CC   contiguous stretch of at least about 60 nucleotides of a sequence
CC   (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences
CC   (GTSs). The novel genes can be used in a process to identify novel
```

```
CC   polynucleotide sequences by comparing them to the novel gene sequences.
CC   The novel genes and cells are useful in functional genomic analysis and
CC   in the discovery and development of new therapeutic and diagnostic agents
CC   and methods. (I) is useful for identifying the coding regions of the
CC   murine genome, to isolate cDNAs, genomic clones or full-length
CC   genes/polynucleotides or homologues, heterologues, paralogues or
CC   orthologues that are capable of hybridising to one or more of the GTSs
CC   under stringent conditions. (I) can be incorporated into a phage display
CC   system that can be used to screen for proteins or other ligands, that are
CC   capable of binding an amino acid sequence encoded by an oligonucleotide
CC   or polynucleotide sequence in at least one of the GTS sequences. (I) is
CC   useful in arrays, such as gene chips, to identify and characterise
CC   temporal and tissue specific gene expression, to identify the gene of
CC   interest from many sources and for genetic manipulations such as
CC   antisense inhibition and gene targeting. Decreasing the level of
CC   expression of (I) and/or down regulating the activity of peptides or
CC   proteins encoded by (I) is useful for treating development and cell
CC   differentiation disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at sequdata.uspto.gov/sequence.html?DocID=20020081668.
XX
SQ   Sequence 397 BP; 89 A; 76 C; 115 G; 113 T; 4 other;

  Query Match             6.7%;  Score 67;  DB 24;  Length 397;
  Best Local Similarity   65.5%;  Pred. No. 1.3e-07;
  Matches   97;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy        242 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 301
              ||| |||   | | || |  | |  |||  |||| | ||||| | | ||  | ||||| |
Db         32 CTCCCTGTTGTTCCTGTTGCCACTTTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGG 91

Qy        302 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 361
               | |||||  |||| | ||||| | | ||  | ||||| | | |||||  |||| | |||
Db         92 ATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATNGCTGGGTTGCTGGGTTG 151

Qy        362 CTCGCTGGCCTGCTTGCTTGCTCGCTGG 389
              || | | ||  | ||||| | | |||||
Db        152 CTGGATTGCTGGATTGCTGGATTGCTGG 179


RESULT 59
AAL40271/c
ID   AAL40271 standard; DNA; 183 BP.
XX
AC   AAL40271;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 31.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   P-PSDB; AAO21918.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear
XX
PS   Example 2; Fig 3; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 183 BP; 69 A; 36 C; 33 G; 45 T; 0 other;

  Query Match             6.7%;  Score 66.8;  DB 24;  Length 183;
  Best Local Similarity   86.0%;  Pred. No. 1.2e-07;
  Matches   74;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        251 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 310
              ||||||||||||| ||| || |||||||||||| ||| || |||||||||||| ||| ||
Db         96 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 37

Qy        311 CTGCTTGCTTGCTCGCTGGCCTGCTT 336
              ||||||||||||| ||| || |||||
Db         36 TTGCTTGCTTGCTTGCTTGCTTGCTT 11


RESULT 60
AAL40264/c
ID   AAL40264 standard; DNA; 186 BP.
XX
AC   AAL40264;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 17.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
```

```
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   P-PSDB; AAO21911.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear -
XX
PS   Example 2; Fig 3; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 186 BP; 71 A; 36 C; 33 G; 46 T; 0 other;

  Query Match             6.7%;  Score 66.8;  DB 24;  Length 186;
  Best Local Similarity   86.0%;  Pred. No. 1.2e-07;
  Matches   74;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        251 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 310
              ||||||||||||| ||| || ||||||||||||| ||| || ||||||||||||| ||| ||
Db         96 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 37

Qy        311 CTGCTTGCTTGCTCGCTGGCCTGCTT 336
              ||||||||||||| ||| || |||||
Db         36 TTGCTTGCTTGCTTGCTTGCTT 11


RESULT 61
AAL40265/c
ID   AAL40265 standard; DNA; 228 BP.
XX
AC   AAL40265;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 19.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   P-PSDB; AAO21912.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear -
XX
PS   Example 2; Fig 3; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 228 BP; 93 A; 42 C; 37 G; 56 T; 0 other;

  Query Match             6.7%;  Score 66.8;  DB 24;  Length 228;
  Best Local Similarity   86.0%;  Pred. No. 1.2e-07;
  Matches   74;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        251 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 310
              ||||||||||||| ||| || ||||||||||||| ||| || ||||||||||||| ||| ||
Db         96 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 37

Qy        311 CTGCTTGCTTGCTCGCTGGCCTGCTT 336
              ||||||||||||| ||| || |||||
Db         36 TTGCTTGCTTGCTTGCTTGCTT 11


RESULT 62
AAL40268/c
ID   AAL40268 standard; DNA; 240 BP.
XX
AC   AAL40268;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 25.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
```

KW Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW auditive nerve damage; delayed speech learning; gene; ds.
XX
OS Haemophilus influenzae.
XX
PN WO200234768-A2.
XX
PD 02-MAY-2002.
XX
PF 24-OCT-2001; 2001WO-EP12390.
XX
PR 24-OCT-2000; 2000GB-0026002.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Thonnard J;
XX
DR WPI; 2002-471431/50.
DR P-PSDB; AAO21915.
XX
PT New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT useful as components of vaccines useful for treating bacterial
PT infection such as otitis media, delayed speech learning, and
PT inflammation of the middle ear -
XX
PS Example 2; Fig 3; 97pp; English.
XX
CC The invention relates to an isolated polypeptide comprising an amino acid
CC sequence which has 85% identity to the entire length of a fully defined
CC non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC in specification. The isolated protein, an immunogenic fragment of the
CC protein or an antibody immunospecific for the protein or immunogenic
CC fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC The molecules may be used in the prevention of adhesion of Gram-negative
CC or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC wounds, to block bacterial adhesion between mammalian extracellular
CC matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC block the normal progression of pathogenesis in infections initiated
CC other than by the implantation of in-dwelling devices. The isolated
CC protein encoding polynucleotide is useful as a research reagent and
CC material for discovery of treatments of and diagnostics for diseases,
CC particularly human diseases. The antibody of the invention is useful for
CC treating infections preferably bacterial infections such as otitis media
CC in infants and children, pneumonia in elders, sinusitis, nosocomial
CC infections, or invasive diseases, chronic otitis media with hearing loss,
CC fluid accumulation in the middle ear, auditive nerve damage, delayed
CC speech learning, infections of the upper respiratory tract and
CC inflammation of the middle ear. This polynucleotide sequence represents a
CC Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ Sequence 240 BP; 98 A; 44 C; 40 G; 58 T; 0 other;

```
  Query Match             6.7%;  Score 66.8;  DB 24;  Length 240;
  Best Local Similarity   86.0%;  Pred. No. 1.3e-07;
  Matches   74;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        251 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 310
              ||||||||||||| ||| || |||||||||||| ||| || |||||||||||| ||| ||
Db         96 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 37

Qy        311 CTGCTTGCTTGCTCGCTGGCCTGCTT 336
              ||||||||||||| ||| || |||||
Db         36 TTGCTTGCTTGCTTGCTTGCTT 11
```

RESULT 63
AAL40266/c
ID AAL40266 standard; DNA; 252 BP.
XX
AC AAL40266;
XX
DT 13-SEP-2002 (first entry)
XX
DE BASB204 related DNA sequence SEQ ID No 21.
XX
KW Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW auditive nerve damage; delayed speech learning; gene; ds.
XX
OS Haemophilus influenzae.
XX
PN WO200234768-A2.
XX
PD 02-MAY-2002.
XX
PF 24-OCT-2001; 2001WO-EP12390.
XX
PR 24-OCT-2000; 2000GB-0026002.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Thonnard J;
XX
DR WPI; 2002-471431/50.
DR P-PSDB; AAO21913.
XX
PT New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT useful as components of vaccines useful for treating bacterial
PT infection such as otitis media, delayed speech learning, and
PT inflammation of the middle ear -
XX
PS Example 1; Fig 3; 97pp; English.
XX
CC The invention relates to an isolated polypeptide comprising an amino acid
CC sequence which has 85% identity to the entire length of a fully defined
CC non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC in specification. The isolated protein, an immunogenic fragment of the
CC protein or an antibody immunospecific for the protein or immunogenic
CC fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC The molecules may be used in the prevention of adhesion of Gram-negative
CC or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC wounds, to block bacterial adhesion between mammalian extracellular
CC matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC block the normal progression of pathogenesis in infections initiated
CC other than by the implantation of in-dwelling devices. The isolated
CC protein encoding polynucleotide is useful as a research reagent and
CC material for discovery of treatments of and diagnostics for diseases,
CC particularly human diseases. The antibody of the invention is useful for
CC treating infections preferably bacterial infections such as otitis media
CC in infants and children, pneumonia in elders, sinusitis, nosocomial
CC infections, or invasive diseases, chronic otitis media with hearing loss,
CC fluid accumulation in the middle ear, auditive nerve damage, delayed
CC speech learning, infections of the upper respiratory tract and
CC inflammation of the middle ear. This polynucleotide sequence represents a
CC Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ Sequence 252 BP; 101 A; 45 C; 45 G; 61 T; 0 other;

```
  Query Match             6.7%;  Score 66.8;  DB 24;  Length 252;
  Best Local Similarity   86.0%;  Pred. No. 1.3e-07;
  Matches   74;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        251 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 310
              ||||||||||||| ||| || |||||||||||| ||| || |||||||||||| ||| ||
Db         96 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 37

Qy        311 CTGCTTGCTTGCTCGCTGGCCTGCTT 336
              ||||||||||||| ||| || |||||
```

```
Db         36 TTGCTTGCTTGCTTGCTTGCTTGCTT 11

RESULT 64
AAL40267/c
ID   AAL40267 standard; DNA; 264 BP.
XX
AC   AAL40267;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 23.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   P-PSDB; AAO21914.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear  -
XX
PS   Example 2; Fig 3; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 264 BP; 110 A; 45 C; 46 G; 63 T; 0 other;

  Query Match             6.7%;  Score 66.8;  DB 24;  Length 264;
  Best Local Similarity   86.0%;  Pred. No. 1.3e-07;
  Matches   74;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        251 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 310
              ||||||||||||| ||| || |||||||||||| ||| || |||||||||||| ||| ||
Db         96 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 37

Qy        311 CTGCTTGCTTGCTCGCTGGCCTGCTT 336
              ||||||||||||| ||| || |||||
Db         36 TTGCTTGCTTGCTTGCTTGCTTGCTT 11


RESULT 65
AAL40269/c
ID   AAL40269 standard; DNA; 267 BP.
XX
AC   AAL40269;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 27.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   P-PSDB; AAO21916.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear  -
XX
PS   Example 2; Fig 3; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
```

```
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 267 BP; 111 A; 45 C; 46 G; 65 T; 0 other;

  Query Match             6.7%;  Score 66.8;  DB 24;  Length 267;
  Best Local Similarity   86.0%;  Pred. No. 1.3e-07;
  Matches   74;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        251 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 310
              ||||||||||||| ||| || |||||||||||| ||| || |||||||||||| ||| ||
Db         96 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 37

Qy        311 CTGCTTGCTTGCTCGCTGGCCTGCTT 336
              ||||||||||||| ||| || |||||
Db         36 TTGCTTGCTTGCTTGCTTGCTTGCTT 11


RESULT 66
AAL40270/c
ID   AAL40270 standard; DNA; 267 BP.
XX
AC   AAL40270;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 29.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   P-PSDB; AAO21917.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear  -
XX
PS   Example 2; Fig 3; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 267 BP; 111 A; 45 C; 46 G; 65 T; 0 other;

  Query Match             6.7%;  Score 66.8;  DB 24;  Length 267;
  Best Local Similarity   86.0%;  Pred. No. 1.3e-07;
  Matches   74;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        251 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 310
              ||||||||||||| ||| || |||||||||||| ||| || |||||||||||| ||| ||
Db         96 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 37

Qy        311 CTGCTTGCTTGCTCGCTGGCCTGCTT 336
              ||||||||||||| ||| || |||||
Db         36 TTGCTTGCTTGCTTGCTTGCTTGCTT 11


RESULT 67
AAS46422
ID   AAS46422 standard; DNA; 13321 BP.
XX
AC   AAS46422;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #144.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW   cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP02955.
XX
PR   15-MAR-2000; 2000DE-1013847.
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumour suppressor
PT   genes and oncogenes, useful in designing primers and probes for
PT   analysing diseases associated with cytosine methylation state e.g.
PT   cancer  -
XX
PS   Claim 1; SEQ ID No 144; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and
CC   oncogenes having a sequence taken from 536 (actually 533 since
CC   numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC   (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC   peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC   form part of a set of probes for detecting the cytosine methylation state
```

```
CC   and/or single nucleotide polymorphisms and also to be used in an
CC   array for analysing diseases associated with CpG dinucleotides e.g.
CC   cancers and tumours. The probes can also be used in a method for
CC   ascertaining genetic and/or epigenetic parameters for the diagnosis
CC   and/or therapy of existing diseases or the predisposition to specific
CC   diseases, by analysing cytosine methylations. The parameters may be
CC   compared to another set of genetic and/or epigenetic parameters, the
CC   differences serving as basis for diagnosis and/or prognosis events which
CC   are disadvantageous to patients. The present sequence is one of the
CC   533 genomic sequences derived from tumour suppressor genes and
CC   oncogenes. Sequences with even numbered Seq ID numbers are the
CC   complementary sequence of the corresponding odd numbered sequence (e.g.
CC   ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC   is missing).
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 13321 BP; 2850 A; 580 C; 3835 G; 6056 T; 0 other;

  Query Match             6.6%;  Score 66.2;  DB 22;  Length 13321;
  Best Local Similarity   45.7%;  Pred. No. 5.9e-07;
  Matches  230;  Conservative    0;  Mismatches  273;  Indels    0;  Gaps    0;

Qy         99 TGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGC 158
              ||   |||     || | | | | | | |||   | | | | ||  ||  |  ||   |
Db      11247 TGGATGGATGAATGGACGGGTGGATGGATGGATGGATGGATAAGTAGATGGACGGGTGGG 11306

Qy        159 TTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGG 218
              | | | || |||  |  ||     | | | | |||   | | | ||   |   ||   |
Db      11307 TGGATAAGTGGATGGATGGATGAGTGGATGGATGGGTGGGTGGATGGATGAGTAGATAGA 11366

Qy        219 CTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 278
                | ||| | | | |  |||  | | |  ||   | |   | | | | |||   | | |
Db      11367 TGGGTGGGTGGATGGATGGGTGGGTGGACGGATGGATGAGTGGATAGATGGGTAGGTGGA 11426

Qy        279 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 338
              | | | | ||    | | | | | | | ||    | | | | | | | | |   | | |
Db      11427 TGGATGGATGAGTGGATGGATGGGTGGATGAATGGATGGATGGGTGGGTAGATGGATGGA 11486

Qy        339 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 398
              | | | | ||| | | |   | | | | |||   | | | | | | | |||   | | |
Db      11487 TGGATGGATGGACGGGTGAGTAGATGGATGGGTGGGTGGATGGATGGATGGGTGGATGGA 11546

Qy        399 TTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGT 458
              |   | | |||   | | | | ||   ||| ||  |||   | | | | ||   |   |
Db      11547 TGAGTGGATGGATGGGTGGGTGGATGGATAGATGGGTGAATGGATGGATAGATGATTGAA 11606

Qy        459 ACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 518
                  |     ||| | |||    || | | |||   | | |  | || | |||   |
Db      11607 TGTATGGGTGAGTGGATGGATGAATGGACGGATGGATGGATGGATGAGCGGATGGATGGG 11666

Qy        519 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGC 578
              | |  ||   | |||     | |  ||   | |||   | | |  ||      ||   |
Db      11667 TGGGTGGATGGATGGATGAGTGGATGGATGGATGGATGGATGGATGAGCGGATGGA 11726

Qy        579 TAGTAGGGCTGCTGGGCTGCTAG 601
              | |  |||  | |||   | |||
Db      11727 TGGATGGGTGGATGGATGGATAG 11749


RESULT 68
ABZ40780/c
ID   ABZ40780 standard; DNA; 147 BP.
XX
AC   ABZ40780;
XX
DT   07-MAR-2003  (first entry)
XX
DE   N. gonorrhoeae nucleotide sequence SEQ ID 6149.
XX
KW   Antibacterial; infection; vaccine; gene therapy; gene; ds.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO200279243-A2.
XX
PD   10-OCT-2002.
XX
PF   12-FEB-2002; 2002WO-IB02069.
XX
PR   12-FEB-2001; 2001GB-0003424.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Fontana MR,  Pizza M,  Masignani V,  Monaci E;
XX
DR   WPI; 2003-058415/05.
DR   P-PSDB; ABP79810.
XX
PT   New protein from Neisseria gonorrheae, useful for the manufacture of a
PT   medicament for treating or preventing N. gonorrheae infection -
XX
PS   Disclosure; Page 637; 815pp; English.
XX
CC   The present invention relates to proteins from Neisseria gonorrhoeae.
CC   Also disclosed are the nucleic acid molecules encoding the proteins and
CC   antibodies that specifically bind to the proteins. The composition
CC   comprising the protein, nucleic acid or antibody is useful for the
CC   manufacture of a medicament for treating or preventing N. gonorrhoeae
CC   infection, this may be in the form of a vaccine or gene therapy.
CC   Sequences given in records ABZ37706-ABZ42016 represent nucleic acid
CC   molecules of the invention.
XX
SQ   Sequence 147 BP; 66 A; 35 C; 31 G; 15 T; 0 other;

  Query Match             6.6%;  Score 66;  DB 25;  Length 147;
  Best Local Similarity   79.6%;  Pred. No. 1.7e-07;
  Matches   78;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy        247 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 306
              | || |||||||||||| ||| || |||||||||||| ||| || |||||||||||| ||
Db        145 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 86

Qy        307 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 344
              | || ||||||||||||| ||| | ||| ||   |  ||
Db         85 TTGCTTGCTTGCTTGCTTGCTTGACTGTTTTTGTTTTC 48


RESULT 69
AAL36820
ID   AAL36820 standard; DNA; 223 BP.
XX
AC   AAL36820;
XX
DT   08-JAN-2002  (first entry)
XX
DE   Human musculoskeletal system related polynucleotide SEQ ID NO 3185.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein;
KW   musculoskeletal system; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155367-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01338.
XX
PR   31-JAN-2000; 2000US-0179065.
```

PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-451937/48.
XX
PT Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders related to the musculoskeletal system including
PT musculoskeletal cancers and also for testing and detection e.g.
PT diagnosis -

```
XX
PS   Example 2; SEQ ID NO 3185; 781pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (AAL34669-AAL37666) and proteins
CC   (ABB03087-ABB04109) associated with the musculoskeletal system useful
CC   for preventing, treating or ameliorating medical conditions e.g. by
CC   protein or gene therapy. The genes are isolated from a range of human
CC   tissues disclosed in the specification. The nucleic acids, proteins,
CC   antibodies and (ant)agonists are useful in the diagnosis, treatment
CC   and prevention of: (a) cancer, e.g. breast and ovarian cancer and
CC   other cancers of the adrenal gland, bone, bone marrow, breast,
CC   gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC   (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC   healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC   and (f) infectious diseases such as viral, bacterial, fungal and
CC   parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 223 BP; 38 A; 35 C; 68 G; 82 T; 0 other;

  Query Match              6.6%;  Score 65.8;  DB 22;  Length 223;
  Best Local Similarity   60.2%;  Pred. No. 2.2e-07;
  Matches  109;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
              ||  | |||    ||  |||||  | |||    ||  |||||| | |||    |   |||
Db         43 CTCTTAGCTCAGTTGGGTGCTTAGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTAGGTGC 102

Qy        299 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 358
              ||| | |||    ||  |||||| | |||    || ||| | ||||  |    ||||||
Db        103 TTGGTAGCTCAGTTGGGTGCTTGGTAGCTCAGGTGTTTGGTAGCTCAGTTAGGTGCTTGG 162

Qy        359 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 418
              | ||||  | |  |||||| | ||||  | |  |||||| | ||||  | |  ||||||
Db        163 TAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGA 222

Qy        419 T 419
              |
Db        223 T 223


RESULT 70
AAK64806
ID   AAK64806 standard; DNA; 223 BP.
XX
AC   AAK64806;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:19618.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
```

PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 19618; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 223 BP; 38 A; 35 C; 68 G; 82 T; 0 other;

Query Match 6.6%; Score 65.8; DB 22; Length 223;
Best Local Similarity 60.2%; Pred. No. 2.2e-07;
Matches 109; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

```
Qy        239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
              ||  | |||    ||  |||||  | |||    ||  ||||||  | |||    |   |||
Db         43 CTCTTAGCTCAGTTGGGTGCTTAGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTAGGTGC 102

Qy        299 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 358
              |||  | |||    ||  ||||||  | |||    ||  |||  | ||||  |    ||||||
Db        103 TTGGTAGCTCAGTTGGGTGCTTGGTAGCTCAGGTGTTTGGTAGCTCAGTTAGGTGCTTGG 162

Qy        359 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 418
              |  ||||  | |  ||||||  | ||||  | |  ||||||  | ||||  | |  ||||||
Db        163 TAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGA 222

Qy        419 T 419
              |
Db        223 T 223
```

RESULT 71
ABX59808
ID ABX59808 standard; cDNA; 223 BP.
XX
AC ABX59808;
XX
DT 26-FEB-2003 (first entry)
XX
DE cDNA encoding novel human musculoskeletal system antigen #2152.
XX
KW Gene; ss; musculoskeletal system antigen; cancer; metastasis;
KW re-vascularisation; thrombosis; arteriosclerosis; mineral content;
KW cardiovascular condition; wound; injury; burn; angiogenesis; ulcer;
KW post-operative tissue repair; limb regeneration; neuronal growth;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW AIDS-related complex; chondrocyte growth; bone regeneration;
KW periodontal regeneration; tissue transport; bone graft; skin aging;
KW keratinocyte growth; hair loss; melanocyte growth; cell proliferation;
KW cell growth; organ transplant; cell differentiation; body height;
KW weight; hair colour; eye colour; skin; percentage of adipose tissue;
KW pigmentation; cosmetic surgery; metabolism; biorhythm; caricadic rhythm;
KW depression; tendency for violence; pain; reproductive capability;
KW hormone level; endocrine level; appetite; libido; memory; stress;
KW storage capability; fat content; lipid content; protein content;
KW carbohydrate content; vitamin content; cofactor content;
KW nutritional component.
XX
OS Homo sapiens.
XX
PN US2002147140-A1.
XX
PD 10-OCT-2002.
XX
PF 17-JAN-2001; 2001US-0764877.
XX

```
PR   31-JAN-2000; 2000US-179065P.
PR   04-FEB-2000; 2000US-180628P.
PR   28-JUN-2000; 2000US-214886P.
PR   07-JUL-2000; 2000US-216647P.
PR   07-JUL-2000; 2000US-216880P.
PR   11-JUL-2000; 2000US-217487P.
PR   11-JUL-2000; 2000US-217496P.
PR   14-JUL-2000; 2000US-218290P.
PR   26-JUL-2000; 2000US-220963P.
PR   26-JUL-2000; 2000US-220964P.
PR   14-AUG-2000; 2000US-224518P.
PR   14-AUG-2000; 2000US-224519P.
PR   14-AUG-2000; 2000US-225267P.
PR   14-AUG-2000; 2000US-225268P.
PR   14-AUG-2000; 2000US-225270P.
PR   14-AUG-2000; 2000US-225447P.
PR   14-AUG-2000; 2000US-225757P.
PR   14-AUG-2000; 2000US-225758P.
PR   22-AUG-2000; 2000US-226868P.
PR   30-AUG-2000; 2000US-228924P.
PR   01-SEP-2000; 2000US-229287P.
PR   01-SEP-2000; 2000US-229343P.
PR   01-SEP-2000; 2000US-229344P.
PR   01-SEP-2000; 2000US-229345P.
PR   05-SEP-2000; 2000US-229509P.
PR   05-SEP-2000; 2000US-229513P.
PR   08-SEP-2000; 2000US-231413P.
PR   21-SEP-2000; 2000US-234223P.
PR   21-SEP-2000; 2000US-234274P.
PR   25-SEP-2000; 2000US-234997P.
PR   27-SEP-2000; 2000US-235834P.
PR   29-SEP-2000; 2000US-236327P.
PR   29-SEP-2000; 2000US-236367P.
PR   29-SEP-2000; 2000US-236368P.
PR   29-SEP-2000; 2000US-236369P.
PR   29-SEP-2000; 2000US-236370P.
PR   02-OCT-2000; 2000US-236802P.
PR   02-OCT-2000; 2000US-237037P.
PR   02-OCT-2000; 2000US-237038P.
PR   02-OCT-2000; 2000US-237039P.
PR   02-OCT-2000; 2000US-237040P.
PR   13-OCT-2000; 2000US-239935P.
PR   20-OCT-2000; 2000US-240960P.
PR   20-OCT-2000; 2000US-241785P.
PR   20-OCT-2000; 2000US-241809P.
PR   01-NOV-2000; 2000US-244617P.
PR   17-NOV-2000; 2000US-249299P.
PR   08-DEC-2000; 2000US-251856P.
PR   08-DEC-2000; 2000US-251868P.
PR   08-DEC-2000; 2000US-251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2003-128199/12.
XX
PT   Isolated nucleic acid molecules encoding musculoskeletal system
PT   associated polypeptides, useful for detecting disorders, e.g. cancer -
XX
PS   Disclosure; SEQ ID NO 3185; 321pp; English.
XX
CC   The invention describes an isolated nucleic acid molecule comprising a
CC   sequence encoding musculoskeletal system associated polypeptides useful
CC   for detecting disorders, e.g., cancer or cancer metastases, in animals
CC   or humans. The nucleic acid: stimulates re-vascularisation of ischaemic
CC   tissues associated with conditions such as thrombosis, arteriosclerosis,
CC   and other cardiovascular conditions; treats wounds due to injuries,
CC   burns, post-operative tissue repair, and ulcers; stimulates angiogenesis
CC   and limb regeneration; stimulates neuronal growth; can treat and prevent
CC   neuronal damage occurring in certain disorders or neurodegenerative
CC   conditions, such as, Alzheimer's disease, Parkinson's disease, and
CC   AIDS-related complex; stimulates chondrocyte growth, thus they can be
CC   used to enhance bone and periodontal regeneration and aid in tissue
CC   transports or bone grafts; prevents skin aging due to sunburn by
CC   stimulating keratinocyte growth; prevents hair loss, since FGF family
CC   members activate hair-forming cells and promotes melanocyte growth;
CC   stimulates growth and differentiation of hematopoietic cells and bone
CC   marrow cells when used in combination with other cytokines; maintains
CC   organs before transplantation or for supporting cell culture of primary
CC   tissues; induces tissue of mesodermal origin to differentiate in early
CC   embryos; increases or decreases the differentiation or proliferation of
CC   embryonic stem cells, besides, haematopoietic lineage; modulates
CC   mammalian characteristics, such as, body height, weight, hair colour, eye
CC   colour, skin, percentage of adipose tissue, pigmentation, size, and shape
CC   (e.g., cosmetic surgery); modulates mammalian metabolism; changes
CC   mammal's metal state or physical state by influencing biorhythms,
CC   caricadic rhythms, depression, tendency for violence, tolerance for pain,
CC   reproductive capabilities, hormonal or endocrine levels, appetite,
CC   libido, memory, or stress; increases or decreases storage capabilities,
CC   fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors
CC   or other nutritional components. This sequence encodes a novel human
CC   musculoskeletal system antigen.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from the US patent office at
CC   ftp.seqdata.uspto.gov/sequence.html?DocID=20020147140.
XX
SQ   Sequence 223 BP; 38 A; 35 C; 68 G; 82 T; 0 other;

  Query Match             6.6%;  Score 65.8;  DB 25;  Length 223;
  Best Local Similarity   60.2%;  Pred. No. 2.2e-07;
  Matches  109;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
              ||  | |||    ||  |||||  | |||    ||  |||||| | |||    |   |||
Db         43 CTCTTAGCTCAGTTGGGTGCTTAGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTAGGTGC 102

Qy        299 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 358
              ||| | |||    ||  ||||||  | |||    || ||| | ||||  |    ||||||
Db        103 TTGGTAGCTCAGTTGGGTGCTTGGTAGCTCAGGTGTTTGGTAGCTCAGTTAGGTGCTTGG 162

Qy        359 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 418
              | ||||  | |  ||||||  | ||||  | |  ||||||  | ||||  | |  ||||||
Db        163 TAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGA 222

Qy        419 T 419
              |
Db        223 T 223


RESULT 72
AAK54984
ID   AAK54984 standard; cDNA; 230 BP.
XX
AC   AAK54984;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:44.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
```

PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
DR P-PSDB; AAM82203.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX

```
PS   Claim 1; SEQ ID NO 44; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 230 BP; 43 A; 36 C; 66 G; 82 T; 3 other;

  Query Match             6.5%;  Score 65.4;  DB 22;  Length 230;
  Best Local Similarity   59.0%;  Pred. No. 2.8e-07;
  Matches  108;  Conservative    1;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
              ||  | |||    ||  |||||  | |||    ||  |||||| | |||    |   :||
Db         43 CTCTTAGCTCAGTTGGGTGCTTAGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTAGGKGC 102

Qy        299 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 358
              ||| | |||  | ||  ||||||  | |||    || ||| | ||||  |    ||||||
Db        103 TTGGTAGCTNACTTGGGTGCTTGGTAGCTCAGGTGTTTGGTAGCTCAGTTAGGTGCTTGG 162

Qy        359 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 418
              | ||||  | |  ||||||  | ||||  | |  ||||||  | ||||  | |  ||||
Db        163 TAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGATAGCTCAGTTGGGTGCTNCA 222

Qy        419 TGA 421
              | |
Db        223 TTA 225


RESULT 73
ABQ97710/c
ID   ABQ97710 standard; cDNA; 393 BP.
XX
AC   ABQ97710;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Mouse ES cell related cDNA SEQ ID NO 978.
XX
KW   Mouse; ES cell; gene trapped sequence; GTS; gene expression;
KW   development disorder; cell differentiation disorder; gene; ss.
XX
OS   Mus sp.
XX
PN   US2002081668-A1.
XX
PD   27-JUN-2002.
XX
PF   30-NOV-2000; 2000US-0728446.
XX
PR   20-NOV-1998;   98US-109302P.
PR   01-DEC-1999;   99US-168270P.
XX
PA   (FRIE/) FRIEDRICH G.
PA   (ZAMB/) ZAMBROWICZ B.
PA   (SAND/) SANDS A T.
XX
PI   Friedrich G,  Zambrowicz B,  Sands AT;
XX
DR   WPI; 2002-626541/67.
XX
PT   Novel murine polynucleotides that individually identify novel genes
PT   into which a retroviral gene trap vector has been integrated, useful in
PT   genomic analysis and in discovery, development of therapeutic and
PT   diagnostic agents -
XX
PS   Claim 2; SEQ ID NO 978; 29pp + Sequence Listing; English.
XX
CC   The invention relates to isolated murine polynucleotides (I) comprising a
CC   contiguous stretch of at least about 60 nucleotides of a sequence
CC   (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences
CC   (GTSs). The novel genes can be used in a process to identify novel
CC   polynucleotide sequences by comparing them to the novel gene sequences.
CC   The novel genes and cells are useful in functional genomic analysis and
CC   in the discovery and development of new therapeutic and diagnostic agents
CC   and methods. (I) is useful for identifying the coding regions of the
CC   murine genome, to isolate cDNAs, genomic clones or full-length
CC   genes/polynucleotides or homologues, heterologues, paralogues or
CC   orthologues that are capable of hybridising to one or more of the GTSs
CC   under stringent conditions. (I) can be incorporated into a phage display
CC   system that can be used to screen for proteins or other ligands, that are
CC   capable of binding an amino acid sequence encoded by an oligonucleotide
CC   or polynucleotide sequence in at least one of the GTS sequences. (I) is
CC   useful in arrays, such as gene chips, to identify and characterise
CC   temporal and tissue specific gene expression, to identify the gene of
CC   interest from many sources and for genetic manipulations such as
CC   antisense inhibition and gene targeting. Decreasing the level of
CC   expression of (I) and/or down regulating the activity of peptides or
CC   proteins encoded by (I) is useful for treating development and cell
CC   differentiation disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at sequdata.uspto.gov/sequence.html?DocID=20020081668.
XX
SQ   Sequence 393 BP; 130 A; 130 C; 66 G; 53 T; 14 other;

  Query Match             6.5%;  Score 65.2;  DB 24;  Length 393;
  Best Local Similarity   54.5%;  Pred. No. 3.7e-07;
  Matches  115;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        226 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 285
              | ||||  | |  |  |  ||||   |  |||||  |   ||  | | ||| | | |
Db        224 CCTGCTGTTGGTCCCTCCTGCTGCATTTNATGCTTTTGCTANTGCTGGATGGCTGGATGG 165

Qy        286 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 345
              ||||  ||||   | ||| | ||||  | ||||| | |  ||||   ||| | ||| |
Db        164 CTGGGTTGCTGNATGGCTGGATGGCTGGGTTGCTGGATGCCTGGATGGCTGGGTTGGTGN 105

Qy        346 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 405
                ||||    |  || | | |||||  ||||   |  || | ||||  |  ||     | |
Db        104 ATGGCTGNATGNCTGGATGGCTGGTTTGCTGNATGNCTGGATGGCTGGGNTGGNCNATGG 45

Qy        406 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 436
              ||||    || | |  | |||  || |  ||
Db         44 CTGGATGNCTGGGTTGGTGGATTGTTGNACT 14


RESULT 74
AAS39915
ID   AAS39915 standard; DNA; 31718 BP.
XX
AC   AAS39915;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Genomic sequence #334 encoding human colon associated polypeptide.
XX
KW   Human; colon cancer; congenital abnormality; infection; colitis;
KW   inflammatory bowel disease; IBD; neoplastic disorder; gene therapy;
KW   intestinal inflammatory disorder; malabsorption syndrome; gastric;
KW   sigmoid disease; antibacterial; antiviral; antiinflammatory;
KW   cytostatic; ds.
```

XX
OS Homo sapiens.
XX
PN WO200155302-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01240.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX

```
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465567/50.
XX
PT   Isolated polypeptide for treating, preventing and/ or prognosing
PT   disorders related to the colon including colon cancers and also for
PT   testing and detection e.g. diagnosis -
XX
PS   Disclosure; SEQ ID No 812; 562pp; English.
XX
CC   The present invention relates to the isolation of novel human colon
CC   associated polypeptides (AAU22468-AAU22701), and the cDNA and genomic
CC   sequences encoding for them. The sequences of the invention are useful
CC   in the diagnosis, treatment, prevention and/or prognosis of disorders
CC   of the colon including colon cancer, congenital abnormalities
CC   (e.g. atresia and stenosis), bacterial and viral infections,
CC   inflammatory bowel disease (IBD), neoplastic cell disorders
CC   (e.g. polyps and adenomas, intestinal inflammatory disorders, colitis,
CC   colonic inflammation, diarrhoea and dysentery, malabsorption syndromes
CC   (e.g. lactose intolerance), intestinal obstruction and sigmoid diseases.
CC   The polynucleotides sequences of the invention can also be used in gene
CC   therapy. AAS39582-AAS40060 represent DNA sequences encoding for the
CC   novel human colon associated polypeptides of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 31718 BP; 7951 A; 7687 C; 8287 G; 7793 T; 0 other;

  Query Match             6.5%;  Score 65.2;  DB 22;  Length 31718;
  Best Local Similarity   62.0%;  Pred. No. 1.4e-06;
  Matches  103;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        246 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 305
              ||  | | ||||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy        306 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy        366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              ||       |   | .||| ||      ||  ||  ||| ||  ||
Db      15036 CTTCTTCTTTCTTTCTCTCTCTTTTTCTCTGTCTCTCTCTCTCTCC 15081


RESULT 75
AAS39916
ID   AAS39916 standard; DNA; 31718 BP.
XX
AC   AAS39916;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Genomic sequence #335 encoding human colon associated polypeptide.
XX
KW   Human; colon cancer; congenital abnormality; infection; colitis;
KW   inflammatory bowel disease; IBD; neoplastic disorder; gene therapy;
KW   intestinal inflammatory disorder; malabsorption syndrome; gastric;
KW   sigmoid disease; antibacterial; antiviral; antiinflammatory;
KW   cytostatic; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155302-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01240.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
```

```
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465567/50.
XX
PT   Isolated polypeptide for treating, preventing and/ or prognosing
PT   disorders related to the colon including colon cancers and also for
PT   testing and detection e.g. diagnosis -
XX
PS   Disclosure; SEQ ID No 813; 562pp; English.
XX
CC   The present invention relates to the isolation of novel human colon
CC   associated polypeptides (AAU22468-AAU22701), and the cDNA and genomic
CC   sequences encoding for them. The sequences of the invention are useful
CC   in the diagnosis, treatment, prevention and/or prognosis of disorders
CC   of the colon including colon cancer, congenital abnormalities
CC   (e.g. atresia and stenosis), bacterial and viral infections,
CC   inflammatory bowel disease (IBD), neoplastic cell disorders
CC   (e.g. polyps and adenomas, intestinal inflammatory disorders, colitis,
CC   colonic inflammation, diarrhoea and dysentery, malabsorption syndromes
CC   (e.g. lactose intolerance), intestinal obstruction and sigmoid diseases.
CC   The polynucleotides sequences of the invention can also be used in gene
CC   therapy. AAS39582-AAS40060 represent DNA sequences encoding for the
CC   novel human colon associated polypeptides of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 31718 BP; 7949 A; 7690 C; 8288 G; 7791 T; 0 other;

  Query Match             6.5%;  Score 65.2;  DB 22;  Length 31718;
  Best Local Similarity   62.0%;  Pred. No. 1.4e-06;
  Matches  103;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        246 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 305
              ||  | | ||||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy        306 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
              ||  | | ||| ||| ||  ||  | |.||| ||| |||  |    | ||| ||| ||
Db      14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy        366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              ||       |   |  ||| ||      ||  ||  ||| ||  ||
Db      15036 CTTCTTCTTTCTTTCTCTCTCTTTTTCTCTGTCTCTCTCTCTCTCC 15081


RESULT 76
AAL06415
ID   AAL06415 standard; DNA; 31718 BP.
XX
AC   AAL06415;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 9103.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
```

PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; SEQ ID NO 9103; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 31718 BP; 7951 A; 7687 C; 8287 G; 7793 T; 0 other;

Query Match 6.5%; Score 65.2; DB 22; Length 31718;
Best Local Similarity 62.0%; Pred. No. 1.4e-06;
Matches 103; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy        246 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 305
              ||  | | |||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy        306 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy        366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              ||       |   |  ||| ||     ||  ||  ||| ||  ||
Db      15036 CTTCTTCTTTCTTTCTCTCTCTTTTTCTCTGTCTCTCTCTCTCTCC 15081


RESULT 77
AAL06416
ID   AAL06416 standard; DNA; 31718 BP.
XX
AC   AAL06416;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 9104.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
```

```
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen
PT   is used in preventing, treating or ameliorating a medical condition -
XX
PS   Disclosure; SEQ ID NO 9104; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention.
XX
SQ   Sequence 31718 BP; 7949 A; 7690 C; 8288 G; 7791 T; 0 other;

  Query Match             6.5%;  Score 65.2;  DB 22;  Length 31718;
  Best Local Similarity   62.0%;  Pred. No. 1.4e-06;
  Matches  103;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        246 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 305
              ||  | | ||||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy        306 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy        366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              ||       |   | ||| ||      ||  ||  ||| ||  ||
Db      15036 CTTCTTCTTTCTTTCTCTCTCTTTTTCTCTGTCTCTCTCTCTCTCC 15081


RESULT 78
AAK73104/c
ID   AAK73104 standard; DNA; 31718 BP.
XX
AC   AAK73104;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27916.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
```

```
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 27916; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 31718 BP; 7793 A; 8287 C; 7687 G; 7951 T; 0 other;

  Query Match             6.5%;  Score 65.2;  DB 22;  Length 31718;
  Best Local Similarity   62.0%;  Pred. No. 1.4e-06;
  Matches  103;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        246 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 305
              ||  | | ||||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      16803 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 16744

Qy        306 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      16743 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 16684

Qy        366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              ||       |   |  ||| ||      ||  ||  ||| ||  ||
Db      16683 CTTCTTCTTTCTTTCTCTCTCTTTTTCTCTGTCTCTCTCTCTCTCC 16638


RESULT 79
AAK73120/c
ID   AAK73120 standard; DNA; 31718 BP.
XX
AC   AAK73120;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27932.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
```

PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 27932; 3071pp + Sequence Listing; English.
XX

```
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 31718 BP; 7791 A; 8288 C; 7690 G; 7949 T; 0 other;

  Query Match             6.5%;  Score 65.2;  DB 22;  Length 31718;
  Best Local Similarity   62.0%;  Pred. No. 1.4e-06;
  Matches  103;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        246 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 305
              ||  | | ||||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      16803 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 16744

Qy        306 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      16743 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 16684

Qy        366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              ||       |   |  ||| ||      ||  ||  ||| ||  ||
Db      16683 CTTCTTCTTTCTTTCTCTCTCTTTTTCTCTGTCTCTCTCTCTCTCC 16638


RESULT 80
AAK87442
ID   AAK87442 standard; DNA; 31718 BP.
XX
AC   AAK87442;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:42254.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
```

```
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 42254; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 31718 BP; 7951 A; 7687 C; 8287 G; 7793 T; 0 other;

  Query Match              6.5%;  Score 65.2;  DB 22;  Length 31718;
  Best Local Similarity   62.0%;  Pred. No. 1.4e-06;
  Matches  103;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        246 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 305
              ||  | | ||||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy        306 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
              ||  | | ||| |||.||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy        366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              ||       |   | ||| ||      ||  ||  ||| ||  ||
Db      15036 CTTCTTCTTTCTTTCTCTCTCTTTTTCTCTGTCTCTCTCTCTCTCC 15081


RESULT 81
AAK87443
ID   AAK87443 standard; DNA; 31718 BP.
XX
AC   AAK87443;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:42255.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
```

PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 42255; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 31718 BP; 7949 A; 7690 C; 8288 G; 7791 T; 0 other;

Query Match 6.5%; Score 65.2; DB 22; Length 31718;
Best Local Similarity 62.0%; Pred. No. 1.4e-06;
Matches 103; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy        246 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 305
              ||  | | |||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy        306 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy        366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              ||       |   |  ||| ||      ||  ||  ||| ||  ||
Db      15036 CTTCTTCTTTCTTTCTCTCTCTTTTTCTCTGTCTCTCTCTCTCTCC 15081


RESULT 82
AAK87573/c
ID   AAK87573 standard; DNA; 31718 BP.
XX
AC   AAK87573;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:42385.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
```

```
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 42385; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 31718 BP; 7793 A; 8287 C; 7687 G; 7951 T; 0 other;

  Query Match             6.5%;  Score 65.2;  DB 22;  Length 31718;
  Best Local Similarity   62.0%;  Pred. No. 1.4e-06;
  Matches  103;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        246 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 305
              ||  | | ||||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      16803 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 16744

Qy        306 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      16743 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 16684

Qy        366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              ||       |   |  ||| ||      ||  ||  ||| ||  ||
Db      16683 CTTCTTCTTTCTTTCTCTCTCTTTTTCTCTGTCTCTCTCTCTCTCC 16638

RESULT 83
AAK87592/c
ID   AAK87592 standard; DNA; 31718 BP.
XX
AC   AAK87592;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:42404.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
```

PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 42404; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 31718 BP; 7791 A; 8288 C; 7690 G; 7949 T; 0 other;

Query Match 6.5%; Score 65.2; DB 22; Length 31718;
Best Local Similarity 62.0%; Pred. No. 1.4e-06;
Matches 103; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy        246 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 305
              ||  | | ||||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      16803 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 16744

Qy        306 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      16743 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 16684

Qy        366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              ||       |   | ||| ||      ||  ||  ||| ||  ||
Db      16683 CTTCTTCTTTCTTTCTCTCTCTTTTTCTCTGTCTCTCTCTCTCTCC 16638
```

RESULT 84
AAK90359
ID AAK90359 standard; DNA; 31718 BP.
XX
AC AAK90359;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen genomic sequence SEQ ID NO: 3935.
XX
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;

KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX
OS Homo sapiens.
XX
PN WO200155314-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01324.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.

```
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-502630/55.
XX
PT   Polynucleotides encoding digestive system antigens, useful for
PT   diagnosing, treating, preventing and/or prognozing disorders of the
PT   digestive system, particularly cancer and cancer metastases -
XX
PS   Disclosure; SEQ ID NO 3935; 986pp; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human digestive system antigens. These can be used in the
CC   diagnosis, treatment and prevention of digestive system disorders,
CC   including cancer, Meckel's diverticulum, bacterial or parasitic
CC   infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC   ulcerative colitis. The present sequence is a genomic DNA fragment
CC   encoding a digestive system antigen of the invention.
XX
SQ   Sequence 31718 BP; 7951 A; 7687 C; 8287 G; 7793 T; 0 other;

  Query Match             6.5%;  Score 65.2;  DB 22;  Length 31718;
  Best Local Similarity   62.0%;  Pred. No. 1.4e-06;
  Matches  103;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        246 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 305
              ||  | | ||||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy        306 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTTCTTTCTTTCTTT 15035

Qy        366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
              ||       |   |  ||| ||      ||  ||  ||| ||  ||
Db      15036 CTTCTTCTTTCTTTCTCTCTTTTTCTCTGTCTCTCTCTCTCTCC 15081


RESULT 85
AAK90360
ID   AAK90360 standard; DNA; 31718 BP.
XX
AC   AAK90360;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human digestive system antigen genomic sequence SEQ ID NO: 3936.
XX
KW   Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW   ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW   digestive system disorder; Meckel's diverticulum; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155314-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01324.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
```

PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases -
XX
PS Disclosure; SEQ ID NO 3936; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention.

XX
SQ Sequence 31718 BP; 7949 A; 7690 C; 8288 G; 7791 T; 0 other;

Query Match 6.5%; Score 65.2; DB 22; Length 31718;
Best Local Similarity 62.0%; Pred. No. 1.4e-06;
Matches 103; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 246 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 305
Db 14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTTCTTT 14975

Qy 306 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 365
Db 14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy 366 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 411
Db 15036 CTTCTTCTTTCTTTCTCTCTCTTTTTCTCTGTCTCTCTCTCTCTCC 15081

RESULT 86
ABK83570/c
ID ABK83570 standard; cDNA; 148834 BP.
XX
AC ABK83570;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #141.
XX
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX
PN WO200228999-A2.
XX
PD 11-APR-2002.
XX
PF 03-OCT-2001; 2001WO-US30821.
XX
PR 03-OCT-2000; 2000US-237189P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
DR WPI; 2002-435328/46.
XX
PT Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
PT drug toxicity -
XX
PS Claim 1; SEQ ID No 141; 114pp; English.
XX
CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially

```
CC   chronic) in a tissue, an allergic response in a subject, exposure of a
CC   subject to a pathogen or sterile inflammatory disease, by detecting the
CC   level of expression in a sample of the tissue of gene(s) from Gs, where
CC   the level of expression of the gene is indicative of inflammation;
CC   (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC   an allergic response in a subject, exposure of a subject to a pathogen
CC   or sterile inflammatory disease, by contacting a tissue having
CC   inflammation with an agent that modulates the expression of gene(s)
CC   from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC   modulating GA; M3 is useful for screening an agent capable of modulating
CC   GCA preferably in an inflammation in a tissue; M4 is useful for
CC   detecting an inflammation (especially chronic) in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC   glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC   reperfusion injury, ARDS, adult respiratory distress syndrome,
CC   inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC   periodontal disease; also bacterial infection, viral infection,
CC   parasitic infection, protozoal infection, fungal infection and M5 is
CC   useful for treating one of the above conditions. The present
CC   sequence represents a gene differentially expressed in granulocytes.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 148834 BP; 38942 A; 33672 C; 34501 G; 41719 T; 0 other;

  Query Match             6.5%;  Score 65;  DB 24;  Length 148834;
  Best Local Similarity   60.5%;  Pred. No. 2.4e-06;
  Matches  107;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy        243 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 302
              || ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| |
Db      44954 TCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC 44895

Qy        303 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 362
              |  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| |
Db      44894 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC 44835

Qy        363 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 419
              |  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||
Db      44834 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 44778


RESULT 87
ABA17155/c
ID   ABA17155 standard; DNA; 32249 BP.
XX
AC   ABA17155;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 9486.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
```

PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 9486; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 32249 BP; 7365 A; 9042 C; 8692 G; 7150 T; 0 other;

Query Match 6.4%; Score 64.4; DB 22; Length 32249;
Best Local Similarity 44.7%; Pred. No. 2.2e-06;
Matches 335; Conservative 0; Mismatches 411; Indels 4; Gaps 2;

Qy 5 GGGAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGG 64
Db 13040 GGGAGGAGGCAGTGAGGGAGGGAAGGAAGGAGGCATGGATGAATGCATGGAGGAAAGAAT 12981

Qy 65 GTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 124
Db 12980 GAATGAATGAATGAATGAATGAATGAATGCATGGGTGAATGGATGGATGGGTGCAT 12921

Qy 125 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 184
Db 12920 GGATGCATGGAGGAATGGATGGATGGATGGATGCATGCATGGAGAAATGGATGGATGGGT 12861

Qy 185 GCTCGCTGGCCTGCTTGCTGAG-GGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCT 243
Db 12860 GCATGGAGGAATGGATGGAGGGATGCACGGAAGAATGGGCAGATGGATGGATGGATGGAT 12801

Qy 244 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 303
Db 12800 GGAAGAATCAATGGAGGGAGTGATGGATAGATGGATTGATGGATGGATGGATGGATGGAT 12741

Qy 304 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 363
Db 12740 GGATGGACGAGTGGATGGATGGGTAAATGGATGGATGGATGAGTGGATGGATGAATGGAT 12681

Qy 364 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGG 423
Db 12680 GGATGGATAGATGGATGCATGGAGGAATGGATGCATGGAGGAATGGATGGATGAGTGGAT 12621

Qy 424 GGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTA 483
Db 12620 GGATAGAGGGAAGAATGGATGAATGAGTGGATGGATAGAGGGAAGAATGGATGGAAGGAA 12561

Qy 484 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGG---GCTGCTAGTAGGGCTGCTGGGCTGCTA 540
Db 12560 GAATTGATGGGTAGATGGATGGATGGATGGATAGATGGGTGAGTGGATGGATGGATGGAT 12501

Qy 541 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 600
Db 12500 GGATGGATGGATGGATGGAAGGATGCATGGAGAAATGGATGGATGGGTGGATGGATGCAT 12441

Qy 601 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 660
Db 12440 GGAGGAATAGATGGATGGATGGATGGATGGATGAATGGATGGAAGGATGGATGCATGCAT 12381

Qy 661 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGG 720
Db 12380 GGAGAAATGGATGGGTGGGTGCATGGAGGAATGGATGGAGGGATGCATGGAAGAATGGGC 12321

```
Qy        721 GGCCGCTTGCCTGCTACTAGGGCTGCTGTG 750
              |  |  ||  ||  |  |  | ||  | |
Db      12320 AGATGGATGGATGGAAGAATCGATGGAGGG 12291


RESULT 88
AAL36821
ID   AAL36821 standard; DNA; 223 BP.
XX
AC   AAL36821;
XX
DT   08-JAN-2002  (first entry)
XX
DE   Human musculoskeletal system related polynucleotide SEQ ID NO 3186.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein;
KW   musculoskeletal system; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155367-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01338.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
```

```
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-451937/48.
XX
PT   Isolated polypeptide for treating, preventing and/ or prognosing
PT   disorders related to the musculoskeletal system including
PT   musculoskeletal cancers and also for testing and detection e.g.
PT   diagnosis -
XX
PS   Example 2; SEQ ID NO 3186; 781pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (AAL34669-AAL37666) and proteins
CC   (ABB03087-ABB04109) associated with the musculoskeletal system useful
CC   for preventing, treating or ameliorating medical conditions e.g. by
CC   protein or gene therapy. The genes are isolated from a range of human
CC   tissues disclosed in the specification. The nucleic acids, proteins,
CC   antibodies and (ant)agonists are useful in the diagnosis, treatment
CC   and prevention of: (a) cancer, e.g. breast and ovarian cancer and
CC   other cancers of the adrenal gland, bone, bone marrow, breast,
CC   gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC   (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC   healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC   and (f) infectious diseases such as viral, bacterial, fungal and
CC   parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 223 BP; 38 A; 34 C; 68 G; 83 T; 0 other;

  Query Match             6.4%;  Score 64.2;  DB 22;  Length 223;
  Best Local Similarity   59.7%;  Pred. No. 5.7e-07;
  Matches  108;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
              ||  | |||    ||  |||||  | |||    ||  |||||| | |||    |   |||
Db         43 CTCTTAGCTCAGTTGGGTGCTTAGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTAGGTGC 102

Qy        299 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 358
              ||| | |||    ||  |||||| | |||    || ||| | ||||  |    || |||
Db        103 TTGGTAGCTCAGTTGGGTGCTTGGTAGCTCAGGTGTTTGGTAGCTCAGTTAGGTGTTTGG 162

Qy        359 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 418
              | ||||  | |  ||||||  | ||||  | |  ||||||  | ||||  | |  ||||||
Db        163 TAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGA 222

Qy        419 T 419
              |
Db        223 T 223

RESULT 89
AAK64807
ID   AAK64807 standard; DNA; 223 BP.
XX
AC   AAK64807;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:19619.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
```

```
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 19619; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 223 BP; 38 A; 34 C; 68 G; 83 T; 0 other;

  Query Match             6.4%;  Score 64.2;  DB 22;  Length 223;
  Best Local Similarity   59.7%;  Pred. No. 5.7e-07;
  Matches  108;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
              ||  | |||    ||  |||||  | |||    ||  |||||| | |||    |  |||
Db         43 CTCTTAGCTCAGTTGGGTGCTTAGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTAGGTGC 102

Qy        299 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 358
              ||| | |||    ||  |||||| | |||    || ||| | ||||  |    || |||
Db        103 TTGGTAGCTCAGTTGGGTGCTTGGTAGCTCAGGTGTTTGGTAGCTCAGTTAGGTGTTTGG 162

Qy        359 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 418
              | ||||  | |  |||||| | ||||  | |  |||||| | ||||  | |  ||||||
Db        163 TAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGA 222

Qy        419 T 419
              |
Db        223 T 223


RESULT 90
ABX59809
ID   ABX59809 standard; cDNA; 223 BP.
```

XX
AC ABX59809;
XX
DT 26-FEB-2003 (first entry)
XX
DE cDNA encoding novel human musculoskeletal system antigen #2153.
XX
KW Gene; ss; musculoskeletal system antigen; cancer; metastasis;
KW re-vascularisation; thrombosis; arteriosclerosis; mineral content;
KW cardiovascular condition; wound; injury; burn; angiogenesis; ulcer;
KW post-operative tissue repair; limb regeneration; neuronal growth;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW AIDS-related complex; chondrocyte growth; bone regeneration;
KW periodontal regeneration; tissue transport; bone graft; skin aging;
KW keratinocyte growth; hair loss; melanocyte growth; cell proliferation;
KW cell growth; organ transplant; cell differentiation; body height;
KW weight; hair colour; eye colour; skin; percentage of adipose tissue;
KW pigmentation; cosmetic surgery; metabolism; biorhythm; caricadic rhythm;
KW depression; tendency for violence; pain; reproductive capability;
KW hormone level; endocrine level; appetite; libido; memory; stress;
KW storage capability; fat content; lipid content; protein content;
KW carbohydrate content; vitamin content; cofactor content;
KW nutritional component.
XX
OS Homo sapiens.
XX
PN US2002147140-A1.
XX
PD 10-OCT-2002.
XX
PF 17-JAN-2001; 2001US-0764877.
XX
PR 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 28-JUN-2000; 2000US-214886P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 22-AUG-2000; 2000US-226868P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 08-SEP-2000; 2000US-231413P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 27-SEP-2000; 2000US-235834P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241809P.
PR 01-NOV-2000; 2000US-244617P.
PR 17-NOV-2000; 2000US-249299P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
XX
PA (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
DR WPI; 2003-128199/12.
XX
PT Isolated nucleic acid molecules encoding musculoskeletal system
PT associated polypeptides, useful for detecting disorders, e.g. cancer -
XX
PS Disclosure; SEQ ID NO 3186; 321pp; English.
XX
CC The invention describes an isolated nucleic acid molecule comprising a
CC sequence encoding musculoskeletal system associated polypeptides useful
CC for detecting disorders, e.g., cancer or cancer metastases, in animals
CC or humans. The nucleic acid: stimulates re-vascularisation of ischaemic
CC tissues associated with conditions such as thrombosis, arteriosclerosis,
CC and other cardiovascular conditions; treats wounds due to injuries,
CC burns, post-operative tissue repair, and ulcers; stimulates angiogenesis
CC and limb regeneration; stimulates neuronal growth; can treat and prevent
CC neuronal damage occurring in certain disorders or neurodegenerative
CC conditions, such as, Alzheimer's disease, Parkinson's disease, and
CC AIDS-related complex; stimulates chondrocyte growth, thus they can be
CC used to enhance bone and periodontal regeneration and aid in tissue
CC transports or bone grafts; prevents skin aging due to sunburn by
CC stimulating keratinocyte growth; prevents hair loss, since FGF family
CC members activate hair-forming cells and promotes melanocyte growth;
CC stimulates growth and differentiation of hematopoietic cells and bone
CC marrow cells when used in combination with other cytokines; maintains
CC organs before transplantation or for supporting cell culture of primary
CC tissues; induces tissue of mesodermal origin to differentiate in early
CC embryos; increases or decreases the differentiation or proliferation of
CC embryonic stem cells, besides, haematopoietic lineage; modulates
CC mammalian characteristics, such as, body height, weight, hair colour, eye
CC colour, skin, percentage of adipose tissue, pigmentation, size, and shape
CC (e.g., cosmetic surgery); modulates mammalian metabolism; changes
CC mammal's metal state or physical state by influencing biorhythms,
CC caricadic rhythms, depression, tendency for violence, tolerance for pain,
CC reproductive capabilities, hormonal or endocrine levels, appetite,
CC libido, memory, or stress; increases or decreases storage capabilities,
CC fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors
CC or other nutritional components. This sequence encodes a novel human
CC musculoskeletal system antigen.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from the US patent office at
CC ftp.seqdata.uspto.gov/sequence.html?DocID=20020147140.
XX
SQ Sequence 223 BP; 38 A; 34 C; 68 G; 83 T; 0 other;

Query Match 6.4%; Score 64.2; DB 25; Length 223;
Best Local Similarity 59.7%; Pred. No. 5.7e-07;
Matches 108; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
|| | ||| || ||||| | ||| || |||||| | ||| | |||
Db 43 CTCTTAGCTCAGTTGGGTGCTTAGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTAGGTGC 102

Qy 299 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 358
||| | ||| || |||||| | ||| || ||| | |||| | || |||
Db 103 TTGGTAGCTCAGTTGGGTGCTTGGTAGCTCAGGTGTTTGGTAGCTCAGTTAGGTGTTTGG 162

```
Qy        359 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 418
              | ||||  | |  ||||||| | ||||  | |  |||||| | ||||  | |  ||||||
Db        163 TAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGA 222

Qy        419 T 419
              |
Db        223 T 223


RESULT 91
AAX08988/c
ID   AAX08988 standard; DNA; 3595 BP.
XX
AC   AAX08988;
XX
DT   12-MAY-1999  (first entry)
XX
DE   Sequence encoding human cerberus like protein.
XX
KW   Cerberus like protein; neuron; neural cell; Schwann cell;
KW   glial cell; astrocyte; precursor cell; stem cell; wound healing;
KW   Parkinsons disease; Alzheimers disease; Lou Gehrig's disease;
KW   osteoporosis; rheumatoid arthritis; osteoarthritis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   exon            117..623
FT                   /*tag=  a
FT   intron          624..2402
FT                   /*tag=  b
FT   exon            2403..2699
FT                   /*tag=  c
XX
PN   WO9901553-A1.
XX
PD   14-JAN-1999.
XX
PF   03-JUN-1998;   98WO-US11462.
XX
PR   03-JUL-1997;   97US-0887997.
XX
PA   (GEMY ) GENETICS INST INC.
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Derobertis EM,  Follettie M;
XX
DR   WPI; 1999-106054/09.
XX
PT   Human and murine cerebus-like proteins - used for treating tissue
PT   defects and degenerative nerve conditions
XX
PS   Disclosure; Page 39-41; 50pp; English.
XX
CC   Compositions containing cerberus like proteins can be used for
CC   inducing the formation of neurons and related neural cells and
CC   tissues, such as Schwann cells, glial cells, and astrocytes, as
CC   well as liver, pancreas, lung, heart, kidney, spleen, stomach, and
CC   cardiac tissue and cells. They may also be used to treat precursor
CC   or stem cells. The compositions can also be used for treating
CC   tissue defects, and healing and maintenance of various types of
CC   tissues and wounds. The cerberus protein containing compositions may
CC   also be used to treat or prevent degenerate nerve conditions such
CC   as Parkinson's disease, Alzheimer's disease, and Lou Gehrig's
CC   disease. They can also be used to treat osteoporosis, rheumatoid
CC   arthritis, osteoarthritis, and other abnormalities of connective
CC   tissue, or of other organs or tissues.
XX
SQ   Sequence 3595 BP; 1110 A; 690 C; 769 G; 1025 T; 1 other;

  Query Match             6.3%;  Score 63.4;  DB 20;  Length 3595;
  Best Local Similarity   61.9%;  Pred. No. 2.1e-06;
  Matches  117;  Conservative    0;  Mismatches   71;  Indels    1;  Gaps    1;

Qy        222 CTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 281
              ||  | | | | ||    |   | ||  |||| | ||| |||  ||  ||||||| ||| ||
Db       3303 CTTTCCTACAAATATTTCACGTAGCCTGCCTCCCTGCCTGCCTGCGTGCCTGCCTGCCTG 3244

Qy        282 CTCGCTGGCCTGCTTG-CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 340
              |  ||  |||||| || ||| ||  ||  | | ||| ||| ||  ||  | | ||| |||
Db       3243 CCTGCCTGCCTGCCTGCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTT 3184

Qy        341 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
              ||  ||  | |  || ||| ||  ||  | | ||| ||  ||| ||  | |||||  ||
Db       3183 CCTTCCTTCCTTCTTTCCTTCCTTCCTTCCTTCCTTCCTCCCTCTCTATCTTGCTTCTTT 3124

Qy        401 GCTCGCTGG 409
              ||| |||||
Db       3123 GCTTGCTGG 3115


RESULT 92
ABX12095/c
ID   ABX12095 standard; DNA; 3595 BP.
XX
AC   ABX12095;
XX
DT   13-MAY-2003  (first entry)
XX
DE   Human cerberus-like protein gene.
XX
KW   Human; cerberus; ds; gene; cerberus-like protein; neuron formation;
KW   neural cell formation; Schwann cell formation; glial cell formation;
KW   liver tissue formation; cardiac tissue formation; wound treatment;
KW   stem cell differentiation; endoderm cell differentiation; osteoporosis;
KW   embryonic cell development; degenerative nerve condition; tissue repair;
KW   Parkinson's disease; Alzheimer's disease; Lou Gehrig's disease; tumour;
KW   connective tissue abnormality; rheumatoid arthritis; osteoarthritis;
KW   gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             117..2699
FT                   /*tag=  a
FT                   /product= "Cerberus-like protein"
FT   exon            117..623
FT                   /*tag=  b
FT                   /number= 1
FT   intron          624..2402
FT                   /*tag=  c
FT                   /number= 1
FT   exon            2403..2699
FT                   /*tag=  d
FT                   /number= 2
XX
PN   US2002164682-A1.
XX
PD   07-NOV-2002.
XX
PF   03-JUN-1998;   98US-0089818.
XX
PR   03-JUL-1997;   97US-0887997.
XX
PA   (FOLL/) FOLLETTIE M.
PA   (DERO/) DEROBERTIS E M.
XX
PI   Follettie M,  Derobertis EM;
XX
DR   WPI; 2003-298696/29.
DR   P-PSDB; ABG76094.
XX
PT   New DNA sequence encoding cerberus protein, useful for treating
PT   conditions involving degradation of, or damage to tissues -
XX
PS   Disclosure; Page 15-17; 21pp; English.
```

```
XX
CC   The invention relates to nucleic acid encoding a murine or human
CC   cerberus-like protein. The protein is useful for the formation of neurons
CC   and related neural cells e.g. Schwann cells and glial cells. The protein
CC   is also useful for the formation of other tissues e.g. liver and cardiac
CC   tissue, for causing stem cells e.g. endoderm cells to differentiate, for
CC   regulation of embryonic cell development. The protein is also useful for
CC   treating wounds by repairing tissue or inducing the formation of new
CC   cells, for treating degenerative nerve conditions e.g. Parkinson's
CC   disease, Alzheimer's disease and Lou Gehrig's disease, for treating
CC   connective tissue abnormalities e.g. osteoporosis, rheumatoid arthritis,
CC   osteoarthritis and for treating tumours. The present sequence represents
CC   the human cerberus-like protein gene.
XX
SQ   Sequence 3595 BP; 1110 A; 690 C; 769 G; 1025 T; 1 other;

  Query Match             6.3%;  Score 63.4;  DB 25;  Length 3595;
  Best Local Similarity   61.9%;  Pred. No. 2.1e-06;
  Matches  117;  Conservative    0;  Mismatches   71;  Indels    1;  Gaps    1;

Qy        222 CTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 281
              ||  | | | | ||   |   | ||  |||| | ||| |||  ||  ||||||| ||| ||
Db       3303 CTTTCCTACAAATATTTCACGTAGCCTGCCTCCCTGCCTGCCTGCGTGCCTGCCTGCCTG 3244

Qy        282 CTCGCTGGCCTGCTTG-CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 340
              |  ||  |||||| || ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| 
Db       3243 CCTGCCTGCCTGCCTGCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTT 3184

Qy        341 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 400
               ||  ||  | |  || ||| ||  ||  | | ||| ||  ||| ||  | ||||| ||
Db       3183 CCTTCCTTCCTTCTTTCCTTCCTTCCTTCCTTCCTTCCTCCCTCTCTATCTTGCTTCTTT 3124

Qy        401 GCTCGCTGG 409
              ||| |||||
Db       3123 GCTTGCTGG 3115


RESULT 93
AAL44951
ID   AAL44951 standard; DNA; 8174 BP.
XX
AC   AAL44951;
XX
DT   13-MAY-2002  (first entry)
XX
DE   Equine herpesvirus type 1 immediate early gene.
XX
KW   Equine herpes virus; immediate early; vaccine; EHV-1; virucide;
KW   immunostimulant; gene; ds.
XX
OS   Equine herpesvirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             988..5451
FT                   /*tag=  a
FT                   /product= "immediate early protein"
XX
PN   WO200209750-A2.
XX
PD   07-FEB-2002.
XX
PF   27-JUL-2001; 2001WO-US23612.
XX
PR   27-JUL-2000; 2000US-0626748.
XX
PA   (RESE ) RESEARCH CORP TECHNOLOGIES INC.
XX
PI   O'Callaghan DJ;
XX
DR   WPI; 2002-206153/26.
DR   P-PSDB; AAO14351.
XX
PT   Novel mutant equine herpesvirus type-1 isolates having mutation in
PT   immediate early gene, useful in formulating vaccine compositions for
PT   preventing and treating equine herpesvirus type-1 infections in horses
PT   -
XX
PS   Disclosure; Page 47-51; 68pp; English.
XX
CC   The present invention relates to a mutant equine herpesvirus type-1
CC   (EHV-1) isolate, in particular a replication-competent EHV-1 isolate
CC   comprising a mutation in the immediate-early (IE) gene of the viral
CC   genome. The mutant isolate is useful for stimulating an immune response,
CC   preferably a cell-mediated or humoral immune response, against EHV-1,
CC   and for preventing or inhibiting an EHV-1 infection in a horse. The
CC   present sequence is the EHV-1 immediate early gene.
XX
SQ   Sequence 8174 BP; 1171 A; 2937 C; 2790 G; 1276 T; 0 other;

  Query Match             6.3%;  Score 63.4;  DB 24;  Length 8174;
  Best Local Similarity   50.8%;  Pred. No. 2.7e-06;
  Matches  151;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy        315 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 374
              ||| || || |  ||  || | |   | | | ||| | | ||| | |   | | |   |
Db       5774 TTGTTTACTGGGCGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGG 5833

Qy        375 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGG 434
                | ||| | |  ||  || | |   | | | ||| | | ||| || | ||   |  |||
Db       5834 CGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGG 5893

Qy        435 CTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGG 494
              | | ||| | |    |  || ||   |  ||   |  || ||||| |  ||| || ||||
Db       5894 CGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGG 5953

Qy        495 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 554
              | | | |  |||| | | |   |   |   |   |  ||| || | |  ||| || ||||
Db       5954 CGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGG 6013

Qy        555 CTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCT 611
              | | | |  ||   |     || |   |   |||  ||   || |  |     | |||
Db       6014 CGGGTTGATGGTTCCTGCTCCTCCCCTTCCTGCTCCTCCCCTTCCTGCTCCTCCCCT 6070


RESULT 94
AAL40260/c
ID   AAL40260 standard; DNA; 149 BP.
XX
AC   AAL40260;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 9.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
```

```
DR   WPI; 2002-471431/50.
DR   P-PSDB; AAO21907.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear -
XX
PS   Example 2; Fig 3; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 149 BP; 62 A; 31 C; 26 G; 30 T; 0 other;

  Query Match             6.3%;  Score 62.8;  DB 24;  Length 149;
  Best Local Similarity   85.4%;  Pred. No. 1.2e-06;
  Matches   70;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
              ||||| ||| || |||||||||||| ||| || |||||||||||| ||| || |||||||
Db         92 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 33

Qy        299 TTGCTCGCTGGCCTGCTTGCTT 320
              ||||| ||| || |||||||||
Db         32 TTGCTTGCTTGCTTGCTTGCTT 11


RESULT 95
AAL40261/c
ID   AAL40261 standard; DNA; 149 BP.
XX
AC   AAL40261;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 11.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   P-PSDB; AAO21908.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear -
XX
PS   Example 2; Fig 3; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 149 BP; 62 A; 31 C; 26 G; 30 T; 0 other;

  Query Match             6.3%;  Score 62.8;  DB 24;  Length 149;
  Best Local Similarity   85.4%;  Pred. No. 1.2e-06;
  Matches   70;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
              ||||| ||| || |||||||||||| ||| || |||||||||||| ||| || |||||||
Db         92 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 33

Qy        299 TTGCTCGCTGGCCTGCTTGCTT 320
              ||||| ||| || |||||||||
Db         32 TTGCTTGCTTGCTTGCTTGCTT 11


RESULT 96
AAL40262/c
ID   AAL40262 standard; DNA; 149 BP.
XX
AC   AAL40262;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 13.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
```

OS Haemophilus influenzae.
XX
PN WO200234768-A2.
XX
PD 02-MAY-2002.
XX
PF 24-OCT-2001; 2001WO-EP12390.
XX
PR 24-OCT-2000; 2000GB-0026002.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Thonnard J;
XX
DR WPI; 2002-471431/50.
DR P-PSDB; AAO21909.
XX
PT New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT useful as components of vaccines useful for treating bacterial
PT infection such as otitis media, delayed speech learning, and
PT inflammation of the middle ear -
XX
PS Example 2; Fig 3; 97pp; English.
XX
CC The invention relates to an isolated polypeptide comprising an amino acid
CC sequence which has 85% identity to the entire length of a fully defined
CC non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC in specification. The isolated protein, an immunogenic fragment of the
CC protein or an antibody immunospecific for the protein or immunogenic
CC fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC The molecules may be used in the prevention of adhesion of Gram-negative
CC or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC wounds, to block bacterial adhesion between mammalian extracellular
CC matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC block the normal progression of pathogenesis in infections initiated
CC other than by the implantation of in-dwelling devices. The isolated
CC protein encoding polynucleotide is useful as a research reagent and
CC material for discovery of treatments of and diagnostics for diseases,
CC particularly human diseases. The antibody of the invention is useful for
CC treating infections preferably bacterial infections such as otitis media
CC in infants and children, pneumonia in elders, sinusitis, nosocomial
CC infections, or invasive diseases, chronic otitis media with hearing loss,
CC fluid accumulation in the middle ear, auditive nerve damage, delayed
CC speech learning, infections of the upper respiratory tract and
CC inflammation of the middle ear. This polynucleotide sequence represents a
CC Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ Sequence 149 BP; 62 A; 31 C; 26 G; 30 T; 0 other;

Query Match 6.3%; Score 62.8; DB 24; Length 149;
Best Local Similarity 85.4%; Pred. No. 1.2e-06;
Matches 70; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```
Qy        239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
              ||||| ||| || |||||||||||| ||| || |||||||||||| ||| || |||||||
Db         92 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 33

Qy        299 TTGCTCGCTGGCCTGCTTGCTT 320
              ||||| ||| || |||||||||
Db         32 TTGCTTGCTTGCTTGCTTGCTT 11
```

RESULT 97
AAL40263/c
ID AAL40263 standard; DNA; 149 BP.
XX
AC AAL40263;
XX
DT 13-SEP-2002 (first entry)
XX
DE BASB204 related DNA sequence SEQ ID No 15.
XX
KW Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW auditive nerve damage; delayed speech learning; gene; ds.
XX
OS Haemophilus influenzae.
XX
PN WO200234768-A2.
XX
PD 02-MAY-2002.
XX
PF 24-OCT-2001; 2001WO-EP12390.
XX
PR 24-OCT-2000; 2000GB-0026002.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Thonnard J;
XX
DR WPI; 2002-471431/50.
DR P-PSDB; AAO21910.
XX
PT New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT useful as components of vaccines useful for treating bacterial
PT infection such as otitis media, delayed speech learning, and
PT inflammation of the middle ear -
XX
PS Example 2; Fig 3; 97pp; English.
XX
CC The invention relates to an isolated polypeptide comprising an amino acid
CC sequence which has 85% identity to the entire length of a fully defined
CC non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC in specification. The isolated protein, an immunogenic fragment of the
CC protein or an antibody immunospecific for the protein or immunogenic
CC fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC The molecules may be used in the prevention of adhesion of Gram-negative
CC or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC wounds, to block bacterial adhesion between mammalian extracellular
CC matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC block the normal progression of pathogenesis in infections initiated
CC other than by the implantation of in-dwelling devices. The isolated
CC protein encoding polynucleotide is useful as a research reagent and
CC material for discovery of treatments of and diagnostics for diseases,
CC particularly human diseases. The antibody of the invention is useful for
CC treating infections preferably bacterial infections such as otitis media
CC in infants and children, pneumonia in elders, sinusitis, nosocomial
CC infections, or invasive diseases, chronic otitis media with hearing loss,
CC fluid accumulation in the middle ear, auditive nerve damage, delayed
CC speech learning, infections of the upper respiratory tract and
CC inflammation of the middle ear. This polynucleotide sequence represents a
CC Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ Sequence 149 BP; 62 A; 31 C; 26 G; 30 T; 0 other;

Query Match 6.3%; Score 62.8; DB 24; Length 149;
Best Local Similarity 85.4%; Pred. No. 1.2e-06;
Matches 70; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```
Qy        239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
              ||||| ||| || |||||||||||| ||| || |||||||||||| ||| || |||||||
Db         92 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 33

Qy        299 TTGCTCGCTGGCCTGCTTGCTT 320
              ||||| ||| || |||||||||
Db         32 TTGCTTGCTTGCTTGCTTGCTT 11
```

RESULT 98
AAL40256/c
ID AAL40256 standard; DNA; 2084 BP.
XX

```
AC   AAL40256;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 1.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   N-PSDB; AAO21901, AAO21902.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear  -
XX
PS   Example 1; Fig 1; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 2084 BP; 706 A; 444 C; 400 G; 534 T; 0 other;

  Query Match             6.3%;  Score 62.8;  DB 24;  Length 2084;
  Best Local Similarity   85.4%;  Pred. No. 2.5e-06;
  Matches   70;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
              ||||| ||| || ||||||||||| ||| || ||||||||||| ||| || |||||||
Db         92 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 33

Qy        299 TTGCTCGCTGGCCTGCTTGCTT 320
              ||||| ||| || |||||||||
Db         32 TTGCTTGCTTGCTTGCTTGCTT 11


RESULT 99
AAL40257/c
ID   AAL40257 standard; DNA; 2084 BP.
XX
AC   AAL40257;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 3.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   N-PSDB; AAO21903, AAO21904.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear  -
XX
PS   Example 1; Fig 1; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 2084 BP; 704 A; 444 C; 401 G; 535 T; 0 other;

  Query Match             6.3%;  Score 62.8;  DB 24;  Length 2084;
  Best Local Similarity   85.4%;  Pred. No. 2.5e-06;
  Matches   70;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;
```

```
Qy        239 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 298
              ||||| ||| || |||||||||||| ||| || |||||.|||||| ||| || |||||||
Db         92 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 33

Qy        299 TTGCTCGCTGGCCTGCTTGCTT 320
              ||||| ||| || |||||||||
Db         32 TTGCTTGCTTGCTTGCTTGCTT 11


RESULT 100
ABL32327
ID   ABL32327 standard; DNA; 6109 BP.
XX
AC   ABL32327;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 300.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 300; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 6109 BP; 1078 A; 195 C; 2061 G; 2771 T; 4 other;

  Query Match             6.3%;  Score 62.8;  DB 24;  Length 6109;
  Best Local Similarity   45.3%;  Pred. No. 3.5e-06;
  Matches  268;  Conservative    0;  Mismatches  322;  Indels    2;  Gaps    1;

Qy        105 GACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTG 164
              ||  | | |  ||  ||   |   |  ||  ||   ||  ||  |  ||   | | | ||
Db       3926 GAATGAATGGGTGGGTGGGTGAGTGGATGGGTGGGTAGATGATGGGTGGATGGGTGGGTG 3985

Qy        165 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTG 224
               | |||     |    | | | | | |||   | | | || | |    |  ||   | ||

Db       3986 GGTGGATGAGTGAATGGGTGGATGGATGGGTGGATAGATGGGTGAGTGGGTGGATAGATG 4045

Qy        225 GCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 284
              | |   |    | |  | | | |||   | | | | | | ||    | | | | | |
Db       4046 GGTGAGTGAGTGAGTGGGTGGATGGGTGGGTGGATGGATAGATGTGTGGATGGGTGGGTG 4105

Qy        285 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 344
              | |||   | | | |   | |  ||   | | | |   | | |||   | |   | | |
Db       4106 GATGGATAGATGGGTGAATGGGCGGAAAGATGGGTGAATGGGTGGATGGATGAATGGATA 4165

Qy        345 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 404
                |||     | | | | | | |||   | | | |   | | |||   | | | | | |
Db       4166 AATGGGTAAATGGGTGGGTGGATGGAGAGATGGGTGAATGGGTGGGTAGGTGGATGGATG 4225

Qy        405 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCT 464
              | ||    | |   |||  ||   | | || ||  ||  ||    || |     | |  |
Db       4226 GGTGTGTGGGTGAATGAATGG--TGGATGGGTGTGTGGGTGGGTGGGTGGATAGATGGGT 4283

Qy        465 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 524
              |  ||   |  ||   | | |  | |||| | |||  ||| | |||| || | |  |
Db       4284 GAATGTGTGGATGGGTGGATAGAGGATGGGAGGATAGATGGGTGGATGGGTTGATGGATG 4343

Qy        525 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 584
              |   | |||   | |     |   | |||   | |    |||    | |   | |||  |
Db       4344 GATGGATGGATGGGTGTATAGATGGATGGATGGGTGAATGGGTGGGTAGATGGATAGATG 4403

Qy        585 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 644
              ||  | ||||  | |    ||     ||||  | | |  |||  | ||||  | | |  |
Db       4404 GGTGGATGGGTCGATGTATGGATGGATGGGTAGATGGGTGGGTGGATGGGTGGATGGGTG 4463

Qy        645 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTG 696
              |   | ||||  | | |  ||   | ||||  | | |   ||  |  |  ||
Db       4464 GATGGGTGGGTGGATGGGTGGATGGGTGGGTGGGTGGATAGGTGGATTAATG 4515


Search completed: August 21, 2003, 21:25:02
Job time : 305.859 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 21, 2003, 20:57:58 ; Search time 2074.54 Seconds
(without alignments)
11727.335 Million cell updates/sec

Title: US-09-831-000-1_COPY_132719_133719
Perfect score: 1001
Sequence: 1 cgccgcccggcatggcggtc.....cgggcacccggcccagatct 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	196.2	19.6	656	28	AZ282406	AZ282406 RPCI-23-1
	2	183.4	18.3	759	28	AZ646178	AZ646178 1M0512A15
c	3	158	15.8	821	29	BZ249369	BZ249369 CH230-409
c	4	141	14.1	684	29	BZ515928	BZ515928 BOMRD40TR
	5	139.6	13.9	545	28	AZ836738	AZ836738 2M0131P07
c	6	138.2	13.8	770	28	BZ182690	BZ182690 CH230-434
	7	136.8	13.7	799	14	CB499552	CB499552 ssalsrkc0
	8	135.4	13.5	749	28	BH484177	BH484177 BOGHS47TR
c	9	135	13.5	1053	13	BX333427	BX333427 BX333427
	10	132.2	13.2	540	28	BH276950	BH276950 CH230-48M
	11	130.8	13.1	712	29	BZ519739	BZ519739 BOMRL03TR
c	12	130.6	13.0	845	28	BH541157	BH541157 BOGLR58TR
	13	130.4	13.0	765	29	BZ221738	BZ221738 CH230-372
c	14	128.4	12.8	894	29	CNS03LWH	AL250010 Tetraodon
	15	127.8	12.8	697	28	BH521799	BH521799 BOGXL17TF
	16	127.2	12.7	597	28	BH537125	BH537125 BOHCS81TR
	17	127.2	12.7	743	29	CNS03LX7	AL250036 Tetraodon
c	18	124.4	12.4	640	29	DR37P5T	AL975317 Danio rer
c	19	123.4	12.3	481	28	AZ828483	AZ828483 2M0105A06
c	20	122.4	12.2	1142	11	AK028099	AK028099 Mus muscu
	21	122.2	12.2	599	28	BH486116	BH486116 BOGJZ09TF
	22	122	12.2	1100	9	AL528199	AL528199 AL528199
	23	120.4	12.0	729	28	BZ041698	BZ041698 11r23h12.
	24	120.2	12.0	693	28	AZ650539	AZ650539 1M0520I14
c	25	119.8	12.0	591	28	AZ633613	AZ633613 1M0488O23
c	26	119.8	12.0	659	10	BB610246	BB610246 BB610246
	27	119.8	12.0	729	28	BH284699	BH284699 CH230-49O
	28	118.6	11.8	420	28	AZ782118	AZ782118 2M0022P01
	29	117.2	11.7	549	29	BZ285588	BZ285588 CH230-385
	30	116.2	11.6	711	12	BM950444	BM950444 UI-M-EH0p
	31	114.8	11.5	1201	13	BX426381	BX426381 BX426381
	32	114.6	11.4	726	28	BZ015565	BZ015565 oeh87c12.
	33	114.4	11.4	669	14	CA048466	CA048466 ssalsrkc0
c	34	114.2	11.4	680	28	AZ608625	AZ608625 1M0433A01
	35	114.2	11.4	736	29	BZ236616	BZ236616 CH230-251
	36	113.8	11.4	633	28	BH604586	BH604586 BOHCF82TF
	37	113.4	11.3	740	28	BH328866	BH328866 CH230-93P
	38	113.2	11.3	712	14	CA047817	CA047817 ssalbrh01
c	39	110.8	11.1	465	28	AZ800556	AZ800556 2M0058F10
	40	110.6	11.0	564	28	BH515818	BH515818 BOHAC43TR
	41	109	10.9	490	28	BH443125	BH443125 BOHRJ25TR
	42	108.8	10.9	462	29	CNS04GRC	AL290001 Tetraodon
c	43	108.4	10.8	660	28	BH474846	BH474846 BOHLI64TF
c	44	108.4	10.8	732	28	BZ060388	BZ060388 11g98b12.
	45	107.4	10.7	821	28	BH725237	BH725237 BOMHG42TF
	46	106.8	10.7	438	9	AI676341	AI676341 etmEST008
	47	106	10.6	711	29	BZ263814	BZ263814 CH230-420
	48	105.4	10.5	528	28	BH700477	BH700477 BOMLB36TF
c	49	105.4	10.5	710	14	CB504454	CB504454 ssalplnb5
	50	105	10.5	569	28	BH483727	BH483727 BOGCW50TF
	51	104.6	10.4	749	28	BH267859	BH267859 CH230-66L
	52	104.4	10.4	794	28	BH544637	BH544637 BOGGJ62TF
c	53	104.2	10.4	1201	13	BX405071	BX405071 BX405071
	54	104	10.4	819	14	CB952830	CB952830 AGENCOURT
	55	102.6	10.2	677	28	BH677755	BH677755 BOHYJ15TR
	56	102.6	10.2	881	29	CNS02WKU	AL217191 Tetraodon
c	57	102.4	10.2	551	29	BX173189	BX173189 Danio rer
	58	102.2	10.2	1200	13	BX447644	BX447644 BX447644
	59	101.8	10.2	705	29	BZ470200	BZ470200 BONQQ85TF
	60	100.6	10.0	416	28	BZ179889	BZ179889 CH230-444
	61	100.2	10.0	457	9	AI676338	AI676338 etmEST008
	62	100.2	10.0	748	29	BZ497497	BZ497497 BONKB78TR
	63	99.6	10.0	1115	13	BX426372	BX426372 BX426372
c	64	99	9.9	702	29	BX242384	BX242384 Danio rer
	65	98.2	9.8	1201	13	BX405071	BX405071 BX405071
c	66	98.2	9.8	2933	11	AK089803	AK089803 Mus muscu
	67	98	9.8	4346	11	AK080786	AK080786 Mus muscu
c	68	97.8	9.8	592	29	BZ216441	BZ216441 CH230-323
c	69	97.4	9.7	566	14	CD347163	CD347163 EtESTef11
c	70	97.2	9.7	394	28	BH704622	BH704622 BOMBE06TR
c	71	97	9.7	787	29	BZ457666	BZ457666 BONRC33TF
c	72	96.8	9.7	477	28	AZ437267	AZ437267 1M0225L23
	73	96.6	9.7	1201	13	BX405072	BX405072 BX405072
	74	96.2	9.6	580	14	CA054386	CA054386 ssalrgb51
	75	96.2	9.6	1094	29	CNS05B34	AL329305 Tetraodon
	76	95.6	9.6	393	10	BG561076	BG561076 EtESTed91
c	77	95.6	9.6	421	28	AZ857141	AZ857141 2M0161O23

c	78	95	9.5	728	29	BX171642	BX171642 Danio rer
c	79	94.8	9.5	310	28	BH115707	BH115707 RPCI-24-2
	80	94.2	9.4	615	9	AI798117	AI798117 wh81d08.x
	81	94	9.4	656	28	AZ282406	AZ282406 RPCI-23-1
c	82	94	9.4	787	28	BH981983	BH981983 ode73b07.
c	83	94	9.4	1115	13	BX426372	BX426372 BX426372
c	84	93.8	9.4	734	14	CA056594	CA056594 ssalrgb53
c	85	93.8	9.4	737	14	CA063770	CA063770 ssalrgb53
	86	93.2	9.3	1039	29	BZ224494	BZ224494 CH230-446
c	87	93	9.3	490	29	DR42N9S	AL981355 Danio rer
c	88	92.8	9.3	1201	13	BX426381	BX426381 BX426381
c	89	92.4	9.2	672	14	CB499018	CB499018 ssalkhc00
	90	92.4	9.2	1152	29	CNS024NL	AL181002 Tetraodon
c	91	92.2	9.2	490	28	BH731712	BH731712 BOMAD94TF
	92	92.2	9.2	661	28	BH325278	BH325278 CH230-45J
c	93	92	9.2	550	29	BX245766	BX245766 Danio rer
c	94	92	9.2	579	14	CD347167	CD347167 EtESTef11
c	95	91.8	9.2	622	28	BH423105	BH423105 BOHJE88TF
	96	91.8	9.2	1200	9	AL562271	AL562271 AL562271
	97	91.4	9.1	404	14	CB811482	CB811482 AMGNNUC:M
c	98	91.4	9.1	745	12	BI688458	BI688458 603316395
c	99	91.2	9.1	461	29	FR0038988	AL126486 Fugu rubr
	100	90.8	9.1	485	28	BH729559	BH729559 BOMAZ10TR
c	101	90.8	9.1	621	28	BZ180072	BZ180072 CH230-302
	102	90.6	9.1	498	14	CB712482	CB712482 AMGNNUC:N
c	103	90.6	9.1	635	10	BB657463	BB657463 BB657463
c	104	90.6	9.1	1673	11	AK084134	AK084134 Mus muscu
c	105	89.4	8.9	673	14	CA052751	CA052751 ssalrgb51
	106	89.4	8.9	923	29	CNS03BUE	AL236975 Tetraodon
c	107	89.2	8.9	612	29	CNS02TXS	AL213769 Tetraodon
	108	88.4	8.8	474	9	AI226677	AI226677 uj13d03.y
	109	88.4	8.8	475	9	AI225652	AI225652 uj13h11.y
	110	88.4	8.8	569	9	AI786538	AI786538 uj17b08.y
	111	88.4	8.8	602	28	AZ798563	AZ798563 2M0055M23
	112	88.4	8.8	604	9	AI317436	AI317436 uj22f12.y
	113	88.2	8.8	526	9	AI786639	AI786639 uj17g02.y
	114	88.2	8.8	622	10	BB651343	BB651343 BB651343
c	115	87.8	8.8	839	29	BZ425593	BZ425593 BONIO62TF
c	116	87.6	8.8	610	29	CNS04LLQ	AL296279 Tetraodon
c	117	87.6	8.8	629	28	AZ364199	AZ364199 1M0110M16
c	118	87.4	8.7	760	28	AZ501999	AZ501999 1M0341N01
c	119	86.6	8.7	778	28	BH645114	BH645114 BOHVI77TR
c	120	86.4	8.6	550	28	AZ500983	AZ500983 1M0339E09
	121	86.4	8.6	629	29	BZ452930	BZ452930 BONFM52TF
	122	86.4	8.6	961	29	CNS041D4	AL270049 Tetraodon
c	123	86.4	8.6	1201	9	AL534303	AL534303 AL534303
	124	86.2	8.6	528	14	CA042084	CA042084 ssalplnb5
	125	85.6	8.6	585	28	BH508316	BH508316 BOGUY28TR
	126	85.6	8.6	1057	29	CNS04CCM	AL284287 Tetraodon
c	127	85.4	8.5	646	28	BH685737	BH685737 BOHXZ76TF
c	128	85.2	8.5	833	29	BZ479906	BZ479906 BONIP76TR
	129	85.2	8.5	1321	29	AG126084	AG126084 Pan trogl
c	130	84.8	8.5	599	28	BH652284	BH652284 BOHVU42TF
c	131	84.8	8.5	711	12	BM950444	BM950444 UI-M-EH0p
c	132	84.6	8.5	833	29	CNS04472	AL273719 Tetraodon
	133	84.4	8.4	508	10	BF459386	BF459386 UI-M-BZ1-
	134	84.4	8.4	519	10	BF459383	BF459383 UI-M-BZ1-
	135	84.4	8.4	675	28	BH420110	BH420110 BOGJL74TF
c	136	84.4	8.4	1201	13	BX405072	BX405072 BX405072
c	137	84.2	8.4	705	14	CA352932	CA352932 624351 NC
c	138	84.2	8.4	923	29	CNS03BUE	AL236975 Tetraodon
c	139	84	8.4	694	28	BH923991	BH923991 odh89f03.
c	140	84	8.4	766	29	CC324930	CC324930 TAM32-2J2
c	141	84	8.4	774	28	BZ143693	BZ143693 CH230-266
c	142	84	8.4	1200	13	BX447644	BX447644 BX447644
	143	83.6	8.4	721	28	BH484515	BH484515 BOHAX34TF
c	144	83.6	8.4	799	14	CB499552	CB499552 ssalsrkc0
c	145	83.2	8.3	327	12	BM305533	BM305533 EtESTee32
	146	83.2	8.3	766	29	CC324930	CC324930 TAM32-2J2
	147	83	8.3	261	28	AZ858629	AZ858629 2M0163G21
c	148	81.6	8.2	849	13	BX462111	BX462111 BX462111
	149	81.4	8.1	407	28	AZ847768	AZ847768 2M0148B03
c	150	81.4	8.1	667	28	AZ707624	AZ707624 RPCI-23-2
c	151	81	8.1	541	9	AI722296	AI722296 fc32b07.x
	152	80.6	8.1	872	29	CNS02859	AL185526 Tetraodon
	153	80.4	8.0	574	29	CNS0243A	AL180271 Tetraodon
c	154	80.4	8.0	713	28	BH547638	BH547638 BOHAC59TF
c	155	80.2	8.0	438	9	AI676341	AI676341 etmEST008
c	156	80.2	8.0	680	10	BG454776	BG454776 NF105F11L
c	157	80	8.0	463	29	BZ217730	BZ217730 CH230-323
c	158	80	8.0	646	28	AZ962905	AZ962905 2M0231G22
	159	80	8.0	941	29	CNS046OU	AL276951 Tetraodon
	160	79.8	8.0	701	28	AZ363103	AZ363103 1M0108B02
c	161	79.6	8.0	354	28	BH515809	BH515809 BOHAC43TF
	162	79.4	7.9	270	28	BH562669	BH562669 BOHPE89TF
c	163	79.4	7.9	1695	29	CC290874	CC290874 CH261-172
	164	79.2	7.9	327	12	BM305533	BM305533 EtESTee32
	165	79.2	7.9	496	10	BG561945	BG561945 EtESTee11
	166	79.2	7.9	511	10	BE028804	BE028804 EtESTea84
	167	79	7.9	598	29	BZ521591	BZ521591 BOKAE54TF
	168	78.8	7.9	212	28	BH248003	BH248003 BOGAC13TF
c	169	78.8	7.9	393	10	BG561076	BG561076 EtESTed91
	170	78.8	7.9	911	29	CNS026IE	AL183407 Tetraodon
	171	78.2	7.8	598	28	BH606983	BH606983 BOHBP37TF
	172	78	7.8	500	13	BX425995	BX425995 BX425995
c	173	77.8	7.8	1101	29	CNS0171Y	AL107488 Drosophil
	174	77.6	7.8	744	28	BZ070693	BZ070693 1kk29f04.
c	175	77.6	7.8	1055	29	CNS05EB2	AL333479 Tetraodon
c	176	77.6	7.8	1200	9	AL562271	AL562271 AL562271
c	177	77.4	7.7	982	13	BX415111	BX415111 BX415111
	178	77.4	7.7	1201	9	AL565450	AL565450 AL565450
	179	77.4	7.7	1297	13	BQ064843	BQ064843 AGENCOURT
	180	77.2	7.7	650	29	BX214340	BX214340 Danio rer
	181	77.2	7.7	1860	11	AK082647	AK082647 Mus muscu
c	182	77	7.7	888	28	BH713235	BH713235 BOMNI36TR
c	183	76.8	7.7	1082	29	CNS05MF7	AL343996 Tetraodon
c	184	76.8	7.7	1274	29	CC209173	CC209173 CH261-44E
	185	76.6	7.7	388	28	BH517322	BH517322 BOGXO85TR
	186	76.6	7.7	579	14	CD347167	CD347167 EtESTef11
	187	76.6	7.7	686	29	BZ511023	BZ511023 BOMRE33TR
c	188	76.4	7.6	420	28	AZ782118	AZ782118 2M0022P01
c	189	76.4	7.6	514	28	AZ251453	AZ251453 RPCI-23-2
c	190	76.4	7.6	545	28	AZ015249	AZ015249 RPCI-23-3
c	191	76.4	7.6	598	29	FR0029401	AL025772 Fugu rubr
c	192	76.2	7.6	719	28	AZ641572	AZ641572 1M0504K12
c	193	76.2	7.6	1005	29	CNS05RJ5	AL350618 Tetraodon
	194	76.2	7.6	1424	13	BQ720687	BQ720687 AGENCOURT
c	195	76	7.6	619	28	AZ422325	AZ422325 1M0200P09
c	196	75.8	7.6	546	29	CNS03QQ0	AL256257 Tetraodon
	197	75.8	7.6	581	14	CD342277	CD342277 EtESTee50
c	198	75.6	7.6	753	28	BH531019	BH531019 BOGYM48TF
c	199	75.6	7.6	827	28	BH250471	BH250471 BOGAK54TR
	200	75	7.5	472	12	BI980396	BI980396 ft75h01.x
	201	75	7.5	743	29	BZ256408	BZ256408 CH230-334
	202	74.8	7.5	510	28	AZ381359	AZ381359 1M0137G22
c	203	74.8	7.5	725	28	AZ859279	AZ859279 2M0164N17
c	204	74.6	7.5	636	12	BM485474	BM485474 pgm1n.pk0
c	205	74.6	7.5	856	29	CNS01X64	AL171301 Tetraodon
	206	74.4	7.4	855	28	BH599667	BH599667 BOGCE95TF
c	207	74.4	7.4	1228	29	CC209174	CC209174 CH261-44E
	208	74.2	7.4	789	29	BZ507343	BZ507343 BONFN66TR
c	209	74.2	7.4	1157	29	CC219063	CC219063 CH261-165
c	210	74	7.4	512	28	BH692899	BH692899 BOMFU72TR
	211	74	7.4	612	28	AZ501249	AZ501249 1M0340H03
	212	74	7.4	654	28	BH430125	BH430125 BOHEN70TF
	213	74	7.4	1057	13	BX349688	BX349688 BX349688
	214	73.8	7.4	585	28	BH550521	BH550521 BOGKO58TR
c	215	73.8	7.4	1110	13	BX456271	BX456271 BX456271
	216	73.6	7.4	261	14	CB498719	CB498719 ssalsrkc0
c	217	73.6	7.4	1066	29	CC206455	CC206455 CH261-138
	218	73.4	7.3	477	13	BX426373	BX426373 BX426373
	219	73.4	7.3	521	28	AZ018805	AZ018805 RPCI-23-2
c	220	73.4	7.3	620	29	CNS048OQ	AL279539 Tetraodon
c	221	73.4	7.3	621	14	CB516309	CB516309 ssalrgb51
	222	73.4	7.3	763	28	AZ705429	AZ705429 RPCI-23-2
	223	73.4	7.3	788	28	AZ643449	AZ643449 1M0506J21

	224	73.4	7.3	850	13	BX404455 BX404455 BX404455 BX404455
	225	73.4	7.3	925	29	CNS0091P AL053013 Drosophil
	226	73.2	7.3	676	29	BZ257915 BZ257915 CH230-479
	227	73.2	7.3	678	29	AG077026 AG077026 Pan trogl
c	228	73.2	7.3	692	28	BH577449 BH577449 BOGTF91TR
	229	73	7.3	592	28	BZ165534 BZ165534 CH230-463
c	230	73	7.3	1767	29	CC294031 CC294031 CH261-82A
	231	72.8	7.3	272	28	BH738439 BH738439 BOMAF71TR
c	232	72.8	7.3	457	9	AI676338 AI676338 etmEST008
c	233	72.8	7.3	693	28	BZ036855 BZ036855 oeh35h07.
c	234	72.8	7.3	935	29	CNS006XK AL066051 Drosophil
	235	72.8	7.3	1201	13	BX340009 BX340009 BX340009
c	236	72.6	7.3	728	28	AZ458530 AZ458530 1M0262O07
	237	72.4	7.2	163	28	AZ764282 AZ764282 1M0560I07
	238	72.4	7.2	635	28	BH699843 BH699843 BOMEH82TR
c	239	72.4	7.2	795	28	BH466067 BH466067 BOGVB34TR
	240	72.4	7.2	866	29	CNS03NVU AL252579 Tetraodon
	241	72.4	7.2	935	29	CNS006XK AL066051 Drosophil
	242	72.4	7.2	1201	13	BX356664 BX356664 BX356664
c	243	72.2	7.2	591	28	AZ611049 AZ611049 1M0436L01
	244	72	7.2	1094	14	CD509997 CD509997 CDA99-D02
c	245	71.8	7.2	216	29	FR0026961 AL019790 F.rubripe
	246	71.8	7.2	1343	12	BM914209 BM914209 AGENCOURT
	247	71.6	7.2	318	9	AW400672 AW400672 LamdiSest
c	248	71.6	7.2	511	10	BE028804 BE028804 EtESTea84
c	249	71.6	7.2	644	29	BZ204163 BZ204163 CH230-351
	250	71.6	7.2	860	29	SSC514813 AJ514813 Sus scrof
	251	71.6	7.2	1004	29	BZ443698 BZ443698 BONDV44TR
	252	71.6	7.2	1056	29	BZ494135 BZ494135 BONKT79TR
	253	71.4	7.1	431	10	BG724542 BG724542 EtESTed88
	254	71.4	7.1	450	10	BG724635 BG724635 EtESTed86
	255	71.4	7.1	487	10	BG516945 BG516945 EtESTed74
c	256	71.4	7.1	771	28	BH701880 BH701880 BOMIF87TF
c	257	71.4	7.1	1063	29	CNS04GZR AL290304 Tetraodon
c	258	71.4	7.1	1491	29	CC236272 CC236272 CH261-190
	259	71.2	7.1	449	28	BZ103827 BZ103827 CH230-147
c	260	71.2	7.1	580	29	DR24J5S AL735176 Danio rer
c	261	71.2	7.1	581	14	CD342277 CD342277 EtESTee50
c	262	71.2	7.1	651	28	AZ314812 AZ314812 1M0031I18
	263	71.2	7.1	696	13	BU608109 BU608109 UI-CF-FN0
c	264	71.2	7.1	755	29	BX158152 BX158152 Danio rer
	265	71.2	7.1	1101	29	CNS0171Y AL107488 Drosophil
	266	71	7.1	326	12	BM102557 BM102557 fu31g01.x
c	267	71	7.1	671	28	AZ989593 AZ989593 2M0273D06
	268	71	7.1	982	13	BX415111 BX415111 BX415111
c	269	71	7.1	1153	29	CC252982 CC252982 CH261-187
c	270	70.8	7.1	620	28	BH304233 BH304233 CH230-10C
c	271	70.8	7.1	1037	29	BZ230139 BZ230139 CH230-527
c	272	70.6	7.1	384	29	CNS043G6 AL272751 Tetraodon
c	273	70.6	7.1	567	29	CNS075AI AL429856 clone BA0
	274	70.4	7.0	239	28	AZ586810 AZ586810 1M0394H05
	275	70.4	7.0	472	29	CC068214 CC068214 CSU-K33r.
	276	70.4	7.0	487	29	CC076917 CC076917 CSU-K33r.
c	277	70.4	7.0	1156	29	CC319460 CC319460 TAM32-17H
	278	70.2	7.0	358	10	BG562249 BG562249 EtESTee04
c	279	70.2	7.0	644	28	BH588140 BH588140 BOHRL57TF
c	280	70.2	7.0	674	28	AZ990550 AZ990550 2M0274M15
	281	70.2	7.0	686	28	BH530855 BH530855 BOHEJ19TR
	282	70.2	7.0	1000	13	BX407619 BX407619 BX407619
c	283	70	7.0	498	28	AZ633443 AZ633443 1M0488N09
c	284	70	7.0	500	13	BX425995 BX425995 BX425995
	285	70	7.0	738	28	BZ078335 BZ078335 llb99e01.
	286	70	7.0	831	12	BI952198 BI952198 HVSMEm000
c	287	70	7.0	848	29	CNS01M45 AL150454 Anopheles
	288	69.8	7.0	300	10	BG560916 BG560916 EtESTed95
c	289	69.8	7.0	576	12	BM582734 BM582734 170006872
c	290	69.8	7.0	616	28	AZ406503 AZ406503 1M0175G13
	291	69.8	7.0	723	29	BX175762 BX175762 Danio rer
	292	69.8	7.0	907	14	CD557158 CD557158 AGENCOURT
	293	69.6	7.0	544	2	HSM095603 Bx501886 Homo sapi
c	294	69.6	7.0	908	29	CNS006B4 AL064031 Drosophil
c	295	69.4	6.9	556	28	AZ847698 AZ847698 2M0148E23
	296	69.4	6.9	992	29	CNS02NJX AL205494 Tetraodon
	297	69.2	6.9	692	28	AZ733538 AZ733538 AZ733538 RPCI-24-1
c	298	69.2	6.9	760	28	BZ051971 BZ051971 BZ051971 jnr56e08.
	299	69.2	6.9	850	29	CNS03W36 AL263211 Tetraodon
c	300	69.2	6.9	1158	29	CC251504 CC251504 CC251504 CH261-84H

ALIGNMENTS

```
RESULT 1
AZ282406/c
LOCUS       AZ282406                 656 bp    DNA     linear   GSS 27-JUL-2000
DEFINITION  RPCI-23-138J12.TJ RPCI-23 Mus musculus genomic clone RPCI-23-138J12
            , genomic survey sequence.
ACCESSION   AZ282406
VERSION     AZ282406.1  GI:9524192
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 656)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished
COMMENT     Other_GSSs: RPCI-23-138J12.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 138  row: J  column: 12
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .656
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-138J12"
                     /sex="Female"
                     /lab_host="DH10B"
                     /clone_lib="RPCI-23"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT      205 a    163 c    282 g      6 t
ORIGIN

  Query Match           19.6%;  Score 196.2;  DB 28;  Length 656;
  Best Local Similarity   56.9%;  Pred. No. 2.5e-26;
  Matches  360;  Conservative    0;  Mismatches  273;  Indels    0;  Gaps    0;

Qy        219 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 278
              | |||||||||||   ||| |||||| ||       | |.|   |||||| ||   ||
Db        649 GTCTGCCTGCCTGTCTGCTTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCT 590

Qy        279 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 338
```

```
Db        589 GTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCT 530

Qy        339 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 398

Db        529 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCT 470

Qy        399 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 458

Db        469 GCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGTCT 410

Qy        459 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 518

Db        409 GCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGCCT 350

Qy        519 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTA 578

Db        349 GTCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTGCCT 290

Qy        579 AGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 638

Db        289 GCCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGTCT 230

Qy        639 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 698

Db        229 GCCTGCCTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCAACATGCCTGCCT 170

Qy        699 CCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCT 758

Db        169 GTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 110

Qy        759 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 818

Db        109 GTCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTGTCTGCCTGCCT 50

Qy        819 GCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGG 851

Db         49 GCCTGCCTGCCTGCCTGCCTGACGGATTGTTGG 17
```

```
RESULT 2
AZ646178
LOCUS       AZ646178                759 bp    DNA     linear   GSS 14-DEC-2000
DEFINITION  1M0512A15F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0512A15 F, genomic survey sequence.
ACCESSION   AZ646178
VERSION     AZ646178.1  GI:11776574
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 759)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0512  row: A  column: 15
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 759.
```

```
FEATURES             Location/Qualifiers
     source          1. .759
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0512A15"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      126 a    167 c    177 g    288 t      1 others
ORIGIN

  Query Match             18.3%;  Score 183.4;  DB 28;  Length 759;
  Best Local Similarity.  67.3%;  Pred. No. 5.4e-24;
  Matches  259;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy        221 CTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGC 280

Db        375 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCGTGCTTGCTTGC 434

Qy        281 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGC 340

Db        435 TTGCGTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCGTGCTTGC 494

Qy        341 TGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGC 400

Db        495 TTGCTTGCTTGCTTGCGTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 554

Qy        401 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 460

Db        555 GTGCGTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 614

Qy        461 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 520

Db        615 GTGCTTGCTTGCTTGCTTGCGTGCTTGCTTGCTTGCTTACGTGCTTGCTTGCTTGC 674

Qy        521 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAG 580

Db        675 TTGCTAGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCGTGCTTGCTAGCTTGC 734

Qy        581 GGGACGGTACGCCTGCCTGATGGCT 605

Db        735 TTGCTTGCTTGCTTGCTTGCTTGCT 759
```

```
RESULT 3
BZ249369/c
LOCUS       BZ249369                821 bp    DNA     linear   GSS 12-OCT-2002
DEFINITION  CH230-409C7.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-409C7, genomic survey sequence.
ACCESSION   BZ249369
VERSION     BZ249369.1  GI:23910598
KEYWORDS    GSS.
```

```
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 821)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-409C7.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 409  row: C  column: 7
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .821
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-409C7"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      274 a    172 c    223 g    152 t
ORIGIN

  Query Match             15.8%;  Score 158;  DB 29;  Length 821;
  Best Local Similarity   64.5%;  Pred. No. 2.3e-19;
  Matches  236;  Conservative    0;  Mismatches  130;  Indels    0;  Gaps    0;

Qy        191 GTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTG 250
              || ||| ||| ||    |     |  ||  | |||  |||||| | ||  || ||| |  ||
Db        394 GTTTGCCTGCCTGCCTGCTTGCCTGTTTGCCTGTTTGCCTGTTTGCCTGCTTGCGTCTTG 335

Qy        251 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 310
                  |       |      | || | | ||||||||||      | | |   || |||||
Db        334 CTTGCTTGCCTGCTTGCTTGCCTTGTTTGCCTGCTTGCCTGCTTGCCTGCTTGCTTGCTT 275

Qy        311 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 370
              |      |   |  ||||||||||||||| ||||||||||     | | |   |  |
Db        274 GTCTGCTTGCTCGCTTGCCTGCTTGCTCGCTTGCCTGCTTGCTTGCCTGCCTGCCTGCTT 215

Qy        371 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 430
              ||| ||||||  |   |  |||| ||| || ||| |||||||  ||| ||||||||| |
Db        214 GCTCGCTTGCCTGCTTGCTTGCTTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCT 155

Qy        431 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              ||| ||| || ||| ||||||  ||  |||||||||| |||||  ||  || |||||
Db        154 GCTTGCTCGCTTGCCTGCTTGCCTGCCTGCCTGCTTGCTCGCTCGCCTGCTCGCCTGCTT 95

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              ||| ||| |||||||||| |||| ||  || ||||||||  |||||| || |||| |||
Db         94 GCTTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCTTGCTCACTCGCTTGCTTGCTCGCTC 35

Qy        551 AGGGGA 556
```

```
              |   ||
Db         34 ATTTGA 29


RESULT 4
BZ515928/c
LOCUS       BZ515928                684 bp  DNA     linear   GSS 16-DEC-2002
DEFINITION  BOMRD40TR BO_2_3_KB Brassica oleracea genomic clone BOMRD40,
            genomic survey sequence.
ACCESSION   BZ515928
VERSION     BZ515928.1  GI:27045003
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 684)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .684
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOMRD40"
                     /clone_lib="BO_2_3_KB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      126 a    237 c    247 g     74 t
ORIGIN

  Query Match             14.1%;  Score 141;  DB 29;  Length 684;
  Best Local Similarity   53.4%;  Pred. No. 2.9e-16;
  Matches  342;  Conservative    0;  Mismatches  295;  Indels    4;  Gaps    2;

Qy        220 GCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGG 279
              || || ||| | |||||   | || ||||  |  |  |    || ||| || |||||
Db        650 GCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGC-TGGTCGCTAC 592

Qy        280 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCG 339
              | || ||||  |  |  |    || ||||  | |        |||   ||  | | |||
Db        591 GCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTAC 532

Qy        340 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGG 399
                || |||| | |       |||   | | | | |||    ||   | | | |||||
Db        531 GCAGCGTGCTCGGTCGCTACGCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTAC 472

Qy        400 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 459
              | || |||  | |||||   | || |||  | |||||   | || |||  | |||||
Db        471 GCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTAC 412

Qy        460 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 519
              | || |||  | |||||   | || |||  | |||||   | || |||  | |||||
Db        411 GCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTAT 352

Qy        520 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAA 579
              | || |||  | |||||  | || |||   |  |  |    |  |||  | | ||||
Db        351 GCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTACGCAGCATGCCCGGTCGCTAC 292

Qy        580 GGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 639
              |  |  |     | || || | | | | ||    |  |||| ||  ||||   |  ||
```

```
Db        291 GCAGCATG---CTCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTG 235

Qy        640 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTC 699
              || ||  ||||    |  |||| ||  ||||    |  |||| ||  ||||    |  |
Db        234 CTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGAGTG 175

Qy        700 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTG 759
              || ||  ||||    |  |||| ||  ||||    |  | || ||  ||||    |  ||
Db        174 CTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTG 115

Qy        760 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 819
              || ||  ||||    |  |||| ||  ||||    |  |||| ||  ||||    |  ||
Db        114 CTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTG 55

Qy        820 CCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGC 860
              |  | | |||     || |||| | | | |  || || |||
Db         54 CTCGGTCGCTACGCAGCGTGCTCGGTCGCGACGCAGCGTGC 14


RESULT 5
AZ836738
LOCUS       AZ836738                545 bp    DNA     linear   GSS 20-FEB-2001
DEFINITION  2M0131P07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0131P07 R, genomic survey sequence.
ACCESSION   AZ836738
VERSION     AZ836738.1  GI:13006646
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 545)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0131  row: P  column: 07
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 545.
FEATURES             Location/Qualifiers
     source          1. .545
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0131P07"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       76 a    151 c    111 g    207 t
ORIGIN

  Query Match           13.9%;  Score 139.6;  DB 28;  Length 545;
  Best Local Similarity   77.5%;  Pred. No. 5.1e-16;
  Matches  169;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        332 CTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTG 391
              |||||| ||| || |||||||       |   |   | ||||| ||||||| |   |  ||
Db         11 CTTGCTTGCTTGCTTGCTTGCCTGCCTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCTTG 70

Qy        392 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 451
              || ||  ||||||||||||||| ||| || |||||||||||| ||  || ||||||||||
Db         71 CTTGCCTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGCTTGCTTGCTTG 130

Qy        452 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 511
              || ||  || ||||||||||  ||| || |||||||||||| ||| || ||| ||||||
Db        131 CTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGCTTG 190

Qy        512 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              || ||| |||||||||||||| ||| ||||||||||
Db        191 CTTGCTTGCCTGCTTGCTTGCTTGCTTGCCTGCTTGCT 228


RESULT 6
BZ182690/c
LOCUS       BZ182690                770 bp    DNA     linear   GSS 11-OCT-2002
DEFINITION  CH230-434H8.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-434H8, genomic survey sequence.
ACCESSION   BZ182690
VERSION     BZ182690.1  GI:23830009
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 770)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-434H8.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 434  row: H  column: 8
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .770
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
```

```
                     /db_xref="taxon:10116"
                     /clone="CH230-434H8"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      120 a    296 c    141 g    213 t
ORIGIN

  Query Match             13.8%;  Score 138.2;  DB 28;  Length 770;
  Best Local Similarity   52.7%;  Pred. No. 9.3e-16;
  Matches  325;  Conservative    0;  Mismatches  288;  Indels    4;  Gaps    1;

Qy        217 ACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGC 276
              | || |||  ||||| ||   ||  |||  ||           | | ||  |  || ||
Db        653 AGGGGTGCAAGCCTGATCAGAGGGGTGCAAGCCTGATCAGAGGGATGCAAGGCTGATCAG 594

Qy        277 TGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGC 336
               ||  |||   ||         |   ||  || || | || ||   | | |||||
Db        593 AGGGATGCAGTCTTGATCAGAGGGGTGCAAGCCTGATCAGAGGGGTGCAAGCCTGATCAG 534

Qy        337 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGC 396
               |   ||  || || | || ||   | | ||||| |     ||  | | | ||| ||
Db        533 AGGGATGCAAGCCTGATCAGAGGGATGCAAGGCTGATCAGAGGCATGCAAGCCTGATCAG 474

Qy        397 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 456
              ||  |||  || || ||   ||  |||  || || ||   ||  |||  || || ||
Db        473 AGGGATGCAAGCCTGATCAGAGGGGTGCAAGCCTGATCAGAGGGGTGCAAGCCTGATCAG 414

Qy        457 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 516
              ||  |||  || || ||   ||  ||   || || ||   ||  |||  || || ||
Db        413 AGGGGTGCAAGCCTGATCAGAGGGATGAAAGCCTGATCAGAGGGGTGCAAGCGTGATCAG 354

Qy        517 TGGCCTGCTTGCTTGCTCGCTGGCCTGC----TTGCTGAGGGGACAGTAGGGCTGCTTGC 572
              ||  |||  |  || ||   ||  |||     || | || ||   | | ||||| |
Db        353 AGGGATGCAAGGCTGATCAGAGGGGTGCAAGTCTGATCAGAGGGATGCAAGGCTGATCAG 294

Qy        573 TTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGT 632
               |       | | |  |    |  ||  || ||||  |||| |||  | ||| |
Db        293 AGGGGTGCAAGCCTGATCAGAGGGGTGCAAGCCTGATCAGAGGGGTGCAAGCCTGATCAG 234

Qy        633 AGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGT 692
              |||| ||   | ||| |   |||| |||  | ||| |   |||| |||  |||| |
Db        233 AGGGATGAAAGCCTGATCAGAGGGGTGCAAGCCTGATCAGAGGGATGCAAGGCTGATCAG 174

Qy        693 AGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGT 752
              |||| | |  ||||| |   |||| |||  | ||| |   |||| | |  | ||| |
Db        173 AGGGGTGCAAGGCTGATCAGAGGGATGCAAGCCTGATCAGAGGGATGCAAGCCTGATCAG 114

Qy        753 AGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGT 812
              |||| |||  | ||| |   |||| |||  | ||| |   |||| |||  | ||| |
Db        113 AGGGATGCAAGCCTGATCAGAGGGATGCAAGCCTGATCAGAGGGATGCAAGCCTGATCAG 54

Qy        813 AGGGCTGCCTGCTGGCT 829
              |||| |||  ||  | |
Db         53 AGGGATGCAAGCCTGGT 37


RESULT 7
CB499552
LOCUS       CB499552                799 bp    mRNA    linear   EST 16-MAY-2003
DEFINITION  ssalsrkc017039 kidney Salmo salar cDNA, mRNA sequence.
ACCESSION   CB499552
VERSION     CB499552.1  GI:29310778
KEYWORDS    EST.
SOURCE      Salmo salar (Atlantic salmon)
  ORGANISM  Salmo salar
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
REFERENCE   1  (bases 1 to 799)
  AUTHORS   GRASP Consortium, Davidson,W.S., Koop,B.F. and
            http://web.uvic.ca/cbr/grasp.
  TITLE     A survey of Salmo salar transcripts from high complexity cDNA
            libraries
  JOURNAL   Unpublished
COMMENT     Contact: Koop BF
            Centre for Biomedical Research
            University of Victoria
            PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
            Tel: 250 472 4067
            Fax: 250 472 4075
            Email: bkoop@uvic.ca
            Centre for Biomedical Research, University of Victoria cDNA
            preparation and sequencing: Roberto Alberto, Marianne Beetz-Sargent
            , Maura Busby, Peter Hunt, Linda McKinnel, BF Koop. bioinformatics:
            Gordon D Brown.
FEATURES             Location/Qualifiers
     source          1. .799
                     /organism="Salmo salar"
                     /mol_type="mRNA"
                     /strain="McConnell"
                     /db_xref="taxon:8030"
                     /clone_lib="kidney"
                     /note="Vector: pBlueScriptIISK+; Library Creator:  Matthew
                     L Rise ; Atlantic salmon tissue contributors: Carlo Biagi,
                     Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.), Simon
                     Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery (Crofton,
                     B.C.), Rachel Roper (University of Victoria)"
BASE COUNT      142 a    195 c    261 g    201 t
ORIGIN

  Query Match             13.7%;  Score 136.8;  DB 14;  Length 799;
  Best Local Similarity   51.9%;  Pred. No. 1.7e-15;
  Matches  334;  Conservative    0;  Mismatches  307;  Indels    3;  Gaps    1;

Qy        223 GCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 282
              ||  ||   || |||| |  ||| |||||  |       | |   || ||| |||| |
Db        137 GCTGGCTGACTAGCTGACTAGCTGGCTGACTGACTGACTGACTGACTGGCTGGCTGACTA 196

Qy        283 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 342
              |||  ||||  |   ||  | |  ||  ||||  |       |||  || |||  ||
Db        197 GCTGACTGACTGACTAGCTGACTGACTGACTGACTAGCTGACTGGCTGACTGGCTGACTA 256

Qy        343 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCT 402
              ||   || |||||  |    |  |    |||| || |||    ||    ||  |||||
Db        257 GCTGACTGGCTGACTGACTGGCTGACTGGCTGACTGGCTGACTGGCTGACTGACTGGCTG 316

Qy        403 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 462
               || |||  || |||| |  ||| |||  || ||| ||   || |||  |||  |||||
Db        317 ACTGGCTGACTGGCTGACTGGCTAGCTGACTGGCTAGCTGACTGGCTAGCTGACTGGCTG 376

Qy        463 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 522
               || ||| |||  |||||  | | |||  || |||| |  |||  || |||  |||||
Db        377 ACTGGCTAGCTGACTGGCTGGTTGGCTGACTGGCTGACTGGCTGACTGGCTGACTGGCTG 436

Qy        523 GCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGG---GACAGTAGGGCTGCTTGCTTGCTAA 579
              || |||  || |||| |  ||| |||||  |    ||| |   | |||  ||  ||
Db        437 ACTAGCTGACTAGCTGACTAGCTGGCTGACTGACTGACTGACTGACTGACTGGCTG 496

Qy        580 GGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 639
                ||| |     ||| ||||  |  |||  |    ||||   | |||   |  |||||
Db        497 ACTGACTGACTAGCTGACTGACTGACTGACTAGCTGACTGACTGACTGACTGGCTG 556

Qy        640 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTC 699
                  ||||       |||||   | |||   |   | |||   |||||   |   | ||
Db        557 ACTAGCTGACTAGCTGGCTGGCAGACTGACTGACTGACTGACTGGCTGACTGGCTGACTG 616

Qy        700 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTG 759
              | |||   |   | |||   |||||   |   | ||    | |||   |   | |||
Db        617 ACTGACTGACTGGCTGACTGGCTGGCTGGCTGACTGACTGACTGACTGACTGGCTGACTG 676

Qy        760 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 819
              |||||       |||||   | |||   |   | |||   | ||    |   |||||
```

```
Db        677 ACTGGCTGACTAGCTGGCTGGCTGACTGACTGACTGACTGACTAGCTGACTGGCTG 736

Qy        820 CCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTG 863
              ||| | || || || || || | ||| | || |||
Db        737 ACTGGCTGACTGACTGACTGACTGGCTGACTGGCTGACTGACTG 780


RESULT 8
BH484177
LOCUS       BH484177                749 bp    DNA     linear   GSS 13-DEC-2001
DEFINITION  BOGHS47TR BOGH Brassica oleracea genomic clone BOGHS47, genomic
            survey sequence.
ACCESSION   BH484177
VERSION     BH484177.1  GI:17692281
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 749)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOGHS47TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .749
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGHS47"
                     /clone_lib="BOGH"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       96 a    221 c    254 g    178 t
ORIGIN

  Query Match           13.5%;  Score 135.4;  DB 28;  Length 749;
  Best Local Similarity   51.8%;  Pred. No. 3e-15;
  Matches  337;  Conservative    0;  Mismatches  306;  Indels    8;  Gaps    1;

Qy        223 GCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 282
              || ||| | ||||| || |||| | | | | || |||| | ||||
Db         96 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGAAGCGTGCTCGGTCGCTACGAA 155

Qy        283 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 342
              || |||| | | | || || | | |          ||   ||  | | |||   |
Db        156 GCGTGCTCGGTCGCTACGTAGCGTGATCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 215

Qy        343 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCT 402
              || |||| | |         ||   |  | | |||  |||   |   | |||||    |
Db        216 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTT 275

Qy        403 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 462
              || |||| | |||||     ||| |||| | |||||     || |||| | |||||
Db        276 GCGTGCTCGGTCGCTACTTTGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 335

Qy        463 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 522
              || |||| | |||||     || |||| | |||||     || |||| | |||||
Db        336 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 395

Qy        523 GCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGG 582
              || || | | |||||     || ||||  | |  |    |  |||| | | |||| |

Db        396 GCGTGGTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 455

Qy        583 GACGGTACGCCT--------GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAG 634
              |   |  ||          || || | | | | ||    |  |||| ||  ||||
Db        456 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 515

Qy        635 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 694
              |   ||| ||  ||||    |  |||| ||  ||||    |  |||| ||  ||||
Db        516 GCGAGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGGAGCGTGCTCGGTCGCTACGTT 575

Qy        695 GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 754
              | || ||  ||||    |  |||| ||  ||||    |  | || ||  ||||
Db        576 ACATGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 635

Qy        755 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 814
              | |||| ||  ||||    |  |||| ||  ||||     ||||  ||  ||||
Db        636 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAACGTGCTCGGTCGCTACGTA 695

Qy        815 GGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCT 865
              | |||  | | |||   | || |||| | | |    |  || |||  | |
Db        696 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGT 746


RESULT 9
BX333427/c
LOCUS       BX333427                1053 bp    mRNA    linear   EST 01-MAY-2003
DEFINITION  BX333427 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DC025YL15 5-PRIME, mRNA sequence.
ACCESSION   BX333427
VERSION     BX333427.1  GI:30312213
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1053)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 5205.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DC025CF08QP1&cluster=5205.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DC025CF08QP1.
FEATURES             Location/Qualifiers
     source          1. .1053
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DC025YL15"
                     /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      255 a    462 c    300 g     36 t
ORIGIN

  Query Match           13.5%;  Score 135;  DB 13;  Length 1053;
  Best Local Similarity   58.3%;  Pred. No. 3.6e-15;
  Matches  460;  Conservative    0;  Mismatches  310;  Indels   19;  Gaps   12;

Qy        219 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 278
              ||||| ||| |||   |||||| ||| |||| |  | | |   | ||||| ||| ||||
Db        884 GGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTG-GCTGGCTGGCTGGCTGCTGGCTGGCTG 826
```

```
Qy        279 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 338
              ||  ||| ||||   ||    |  | |||  ||    |  | | |   | ||||| |||
Db        825 GCTGGCTGGCTGGCTGGCTGCT--GGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTG 768

Qy        339 GCTGGCCTGCTTGCTGAGGGGACAGTAG--GGCTGCTGGCTTGCTAGTAGGGCTGCTCGC 396
              ||||||  ||| ||||   ||    | |  |||||   |  ||   |   ||||||| ||
Db        767 GCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGC 708

Qy        397 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 456
              ||||  ||| ||| ||| ||||| ||||| ||| ||| ||||||  ||| ||| ||| ||
Db        707 TGGCTGGCTGGCTGGCTGGCTGG-CTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT-GC 650

Qy        457 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 516
              ||||  ||| ||| ||| ||||| ||| |||| ||| ||||||  ||| ||| ||| ||
Db        649 TGGCTGGCTGGCTGGCTGGCTGG-CTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGC 591

Qy        517 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGG--GACAGTAGGGCTGCTTGCTT 574
              | ||  ||| ||| ||| ||||||  ||| ||||  ||  | | |   |||||  ||  |
Db        590 T-GCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCT 532

Qy        575 GCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAG--GGCTGCTGGGCTGCTAGT 632
              |       ||  ||   ||  ||  | ||||| | |  | |  |||||    |||||   |
Db        531 GGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGG 472

Qy        633 AGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG--GGCTGCTGGGCTGCTA 690
                |||||||||   ||| |  ||   |||||   |||  | |  |||||   |||||
Db        471 CTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCT 412

Qy        691 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG--GGCTCCTGGGCTGC 748
              |   |||| ||||   ||| |  ||   |||||   |||  | |  ||||    |||||
Db        411 GGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGG 352

Qy        749 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG--GGCTGCTGGGCT 806
                |  |||||||||   ||| |  ||   |||||   |||  | |  |||||   ||||
Db        351 CTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCT 292

Qy        807 GCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTA 866
              |   |   ||||||  ||||||| ||| ||| ||| |||  |  || | |   |  |||
Db        291 GGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCCTGGGCGGATGCAC 232

Qy        867 CTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGGGCTGCCGGCCAGCTGCGTGCGAGGG 926
                |  ||  ||| |   |      ||   |   |   ||   | | |||    || ||
Db        231 TGGCAGTGGCAATGCTGGCCGTCTGGCAACTTTTCGTCTCCCCGCGGATGCCCTGGACGG 172

Qy        927 CGTCCGAGGGCCAGACGAGGACACGGGACCCGGGCCTCTCCCCCGCCCGGACCGCCGGGC 986
              |   | |  || ||    | |          ||  |  |||| ||  | |   | | || |
Db        171 CCCTCCACTGCGAGGGCACGGACGCCACACCCAGGGTCTCGTCCTGCAGCGGCCCTGGCC 112

Qy        987 ACCCGGCCC 995
                |||  | ||
Db        111 GCCCCGGCC 103
```

```
RESULT 10
BH276950
LOCUS       BH276950                 540 bp    DNA     linear   GSS 30-NOV-2001
DEFINITION  CH230-48M4.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
            CH230-48M4, genomic survey sequence.
ACCESSION   BH276950
VERSION     BH276950.1  GI:17189352
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 EcoRI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-48M4.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 48  row: M  column: 4
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-48M4"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 1"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      112 a    138 c    121 g    169 t
ORIGIN

  Query Match             13.2%;  Score 132.2;  DB 28;  Length 540;
  Best Local Similarity   77.0%;  Pred. No. 1.1e-14;
  Matches  161;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        341 TGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGC 400
              | ||  || |||||  |||   |   || |||| ||||||| |   |  |||| ||| ||
Db        327 TAGCTAGCCTGCTGGTGGGCTTGCTAGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 386

Qy        401 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 460
               ||||||||||||| ||| || |||||||||||| ||| || |||||||||||| ||| ||
Db        387 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 446

Qy        461 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 520
               ||||||||||||| ||| || |||||||||||| ||| || |||||||||||| ||| ||
Db        447 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 506

Qy        521 CTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
               |||||||||||  |||  | | ||| ||
Db        507 TAGCTTGCTTGCTTGCTTCCTTTCTTCCT 535
```

```
RESULT 11
BZ519739
LOCUS       BZ519739                 712 bp    DNA     linear   GSS 16-DEC-2002
DEFINITION  BOMRL03TR BO_2_3_KB Brassica oleracea genomic clone BOMRL03,
            genomic survey sequence.
ACCESSION   BZ519739
VERSION     BZ519739.1  GI:27051945
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magholiophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 712)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Contact: Chris Town
            TIGR
```

```
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .712
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOMRL03"
                     /clone_lib="BO_2_3_KB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       83 a    238 c    243 g    148 t
ORIGIN

  Query Match           13.1%;  Score 130.8;  DB 29;  Length 712;
  Best Local Similarity   50.7%;  Pred. No. 2.1e-14;
  Matches  341;  Conservative    0;  Mismatches  327;  Indels    4;  Gaps    1;

Qy        227 GCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTT 286
              || |||||| | |||   | ||       |   |       || |||||| | |||   |
Db         41 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 100

Qy        287 GCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 346
              ||      |   |       || ||||  |  | ||    || |||| | ||||
Db        101 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 160

Qy        347 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 406
              || ||||  |  | ||    |  |||| | | || |    |  |||||| | |||   |
Db        161 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 220

Qy        407 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 466
              || |||||| | |||   | || |||||| | |||   | || |||||| | |||   |
Db        221 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 280

Qy        467 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 526
              || |||||| | |||   | || |||||| | |||   | || |||||| | |||   |
Db        281 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 340

Qy        527 GCTTGCTCGCTGGCC----TGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGG 582
              || |||||| | |||     || ||||  |  | ||     | |||| | | || | |
Db        341 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 400

Qy        583 GACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 642
              |   |  ||  |||  |||| |||| |   | |  | |  || | |   | |  ||
Db        401 GCGTGATCGGTCGCCACGTGGCGTGATCGGTCGCCACGTAGCGTGATCGGTCGCCACGTG 460

Qy        643 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTG 702
              |  |||| |   | |   | |  || | |   | |   | |  |||| |   | | | |
Db        461 GCGTGCTCGGTCGCCACGTAGCGTGTTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 520

Qy        703 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTG 762
              |  |||| |   | |   | |  |||| |   | | | |  |||| |   | |   |
Db        521 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 580

Qy        763 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCT 822
              |  |||| |   | |   | |  |||| |   | |   | |  |||| |   | |
Db        581 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 640

Qy        823 GCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCA 882
              ||  ||| | | ||    | ||  |   || ||||     || ||   |  | || |  |
Db        641 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCTACGTAGCGACCGTGCTTTCTTTAAA 700

Qy        883 GCTGGGAGAACA 894
              |  |  |||
Db        701 ATCGATACGACA 712
```

```
RESULT 12
BH541157/c
LOCUS       BH541157                845 bp    DNA     linear   GSS 14-DEC-2001
DEFINITION  BOGLR58TR BOGL Brassica oleracea genomic clone BOGLR58, genomic
            survey sequence.
ACCESSION   BH541157
VERSION     BH541157.1  GI:17789521
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 845)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOGLR58TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .845
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGLR58"
                     /clone_lib="BOGL"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      200 a    307 c    254 g     84 t
ORIGIN

  Query Match           13.0%;  Score 130.6;  DB 28;  Length 845;
  Best Local Similarity   51.3%;  Pred. No. 2.3e-14;
  Matches  334;  Conservative    0;  Mismatches  309;  Indels    8;  Gaps    1;

Qy        223 GCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 282
              || |||  | |||||     || ||||  |  |  |     || |||| | |||||
Db        843 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 784

Qy        283 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 342
              || ||||  |  |  |     || |||| | |         ||   ||  | | |||   |
Db        783 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 724

Qy        343 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCT 402
              || |||| | |         ||   |  | | |||   |||   |  | |||||
Db        723 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 664

Qy        403 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 462
              || |||| | |||||     || |||| | |||||     || |||| | |||||
Db        663 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 604

Qy        463 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 522
              || |||| | |||||     || |||| | |||||     || |||| | |||||
Db        603 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 544

Qy        523 GCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGG 582
              || |||| | |||||     || ||||  |  |  |     |  |||| | | |||| |
Db        543 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 484

Qy        583 GACGGTACGCCT--------GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAG 634
              |   |  ||           || || | | | | ||    |  |||| ||  ||||
Db        483 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 424

Qy        635 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 694
```

```
Db        423 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 364

Qy        695 GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 754

Db        363 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCGACGTAGCGTGCTCGGTCGCGACGTA 304

Qy        755 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 814

Db        303 GCGTGCTCGGTCGCGACGTAGCGTGCTCGGTCGCGACGTAGCGTGCTCGGTCGCGACGTA 244

Qy        815 GGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCT 865

Db        243 GCGTGCTCGGTCGCGACGTAGCGTGCTCGGTCGCGACGTAGCGTGCTCGGT 193


RESULT 13
BZ221738
LOCUS       BZ221738                765 bp    DNA     linear   GSS 11-OCT-2002
DEFINITION  CH230-372D22.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-372D22, genomic survey sequence.
ACCESSION   BZ221738
VERSION     BZ221738.1  GI:23880096
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 765)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-372D22.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 372  row: D  column: 22
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .765
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-372D22"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      213 a    137 c    302 g    113 t
ORIGIN

  Query Match            13.0%;  Score 130.4;  DB 29;  Length 765;
  Best Local Similarity  54.4%;  Pred. No. 2.5e-14;
  Matches  286;  Conservative    0;  Mismatches  236;  Indels    4;  Gaps    1;

Qy        303 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 362

Db        239 GCAAGGCTGATCAGAGGGATGCAAGCCTGATCAGAGGGGTGCAAGCCTGATCAGAGGGGT 298

Qy        363 GTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 422

Db        299 ACAAGCCTGATCAGAGGGATGCAAGGCTGATCAGAGGGGTACAAGCCTGATCAGAGGGGT 358

Qy        423 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 482

Db        359 GCAAGGCTGATCAGAGGGATGCAAGCCTGATCAGAGGGATGCAAGCCTGATCAGAGGGAT 418

Qy        483 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 542

Db        419 GCAAGCCTGATCAGAGGGATGCAAGCCTGATCAGAGGGGTACAAGCCTGATCAGAGGGAT 478

Qy        543 GC----TTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCT 598

Db        479 GCAAGGCTGATCAGAGGGGTGCAAGCCTGATCAGAGGGGTGCAAGGCTGATCAGAGGGAT 538

Qy        599 GATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 658

Db        539 GCAAGCCTGATCAGAGGGGTGCAAGGCTGATCAGAGGGATGCAAGGCTGATCAGAGGGGT 598

Qy        659 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCT 718

Db        599 GCAAGCCTGATCAGAGGGGTGCAAGGCTGATCAGAGGGGTGCAAGCCTGATCAGAGGGGT 658

Qy        719 GCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 778

Db        659 GCAAGGCTGATCAGAGGGGTGCAAGCCTGATCAGAGGGGTGCAAGGCTGATCAGAAGGGT 718

Qy        779 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGC 824

Db        719 GCAAGGCTGATCAGAGGGGTGCAAGGCTGATCAAAGGGGTGCAAGC 764


RESULT 14
CNS03LWH/c
LOCUS       CNS03LWH                894 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            036N09 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL250010
VERSION     AL250010.1  GI:7971022
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 894)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
```

```
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .894
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="036N09"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG036CG05SP1~end :
                     PUC-Ori"
BASE COUNT      243 a    396 c    221 g     32 t      2 others
ORIGIN

  Query Match             12.8%;  Score 128.4;  DB 29;  Length 894;
  Best Local Similarity   59.5%;  Pred. No. 5.7e-14;
  Matches  450;  Conservative    0;  Mismatches  286;  Indels   20;  Gaps   13;

Qy        219 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 278
              ||||| ||| |||   ||||||  ||| |||| |  | | |   | |||||  ||| ||||
Db        849 GGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTG-GCTGGCTGGCTGGCTGCTGGCTGGCTG 791

Qy        279 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 338
              ||  ||| ||||   ||    |  | |||  ||    |  | | |   | ||||| |||
Db        790 GCTGGCTGGCTGGCTGGCTGCT--GGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTG 733

Qy        339 GCTGGCCTGCTTGCTGAGGGGACAGTAG--GGCTGCTGGCTTGCTAGTAGGGCTGCTCGC 396
              ||||||  ||| ||||   ||    | |  |||||   |  ||   |   |||||||  ||
Db        732 GCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGC 673

Qy        397 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 456
              ||||  ||| ||| ||| |||||  |||||  ||| |||  ||||||  ||| ||| ||| ||
Db        672 TGGCTGGCTGGCTGGCTGGCTGG-CTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT-GC 615

Qy        457 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 516
              ||||  ||| ||| |||  |||||  ||| |||| ||| ||||||  ||| ||| ||| ||
Db        614 TGGCTGGCTGGCTGGCTGGCTGG-CTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGC 556

Qy        517 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGG--GACAGTAGGGCTGCTTGCTT 574
              | ||  ||| ||| ||| ||||||  ||| ||||  ||  | | |   |||||  ||  |
Db        555 T-GCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 497

Qy        575 GCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAG--GGCTGCTGGGCTGCTAGT 632
              |      ||  ||   ||  ||  | |||||  | |   | |  |||||   |||||   |
Db        496 GGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGG 437

Qy        633 AGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG--GGCTGCTGGGCTGCTA 690
              |||||||||   ||| |  ||   |||||  |||  | |  |||||   |||||
Db        436 CTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCT 377

Qy        691 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG--GGCTCCTGGGCTGC 748
              |  |||| ||||   ||| |  ||   |||||  ||| | |  |||||   |||||
Db        376 GGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGG 317

Qy        749 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 808
              |  |||||||||   ||| |  ||   |||||  ||| |  |||||   |||||
Db        316 CTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCT--GGCTGGCTGGCTGG 259

Qy        809 TAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACT 868
              |  |||||  |||||||||| ||| ||| ||| ||| |   ||| ||| ||  |||
Db        258 CTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCT-GGCTGGCTGCTGGCTGGCTGGC 200

Qy        869 AGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGGGCTGCCGGCCAGCTGCGTGCGAGGGCG 928
              ||  ||||    ||||  |       |  ||||| | | |     ||  ||  |
Db        199 TGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTG 140

Qy        929 TCCGAGGGCCAGACGAGGACACGGGACCCGGGCCTC 964
              | |   | | | |   |     ||     || ||||
Db        139 GCTGGCTGGCTGGCTGGCTGCTGGCTGATGGACCTC 104
```

```
RESULT 15
BH521799
LOCUS       BH521799                697 bp    DNA     linear   GSS 13-DEC-2001
DEFINITION  BOGXL17TF BOGX Brassica oleracea genomic clone BOGXL17, genomic
            survey sequence.
ACCESSION   BH521799
VERSION     BH521799.1  GI:17729884
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 697)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOGXL17TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .697
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGXL17"
                     /clone_lib="BOGX"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       84 a    232 c    231 g    150 t
ORIGIN

  Query Match             12.8%;  Score 127.8;  DB 28;  Length 697;
  Best Local Similarity   51.4%;  Pred. No. 7.3e-14;
  Matches  323;  Conservative    0;  Mismatches  302;  Indels    4;  Gaps    1;

Qy        227 GCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTT 286
              || |||||| | |||   | ||       |   |       || |||||| | |||   |
Db          3 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 62

Qy        287 GCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 346
              ||      |   |       || ||||  |  | ||    || |||| | ||||
Db         63 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 122

Qy        347 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 406
              || ||||  |  | ||    |  |||| | | || |    |  |||||| | |||   |
Db        123 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 182

Qy        407 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 466
              || |||||| | |||   | ||  |||||| | |||   | ||  |||||| | |||   |
Db        183 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 242

Qy        467 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 526
              || |||||| | |||   | ||  |||||| | |||   | ||  |||||| | |||   |
Db        243 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 302

Qy        527 GCTTGCTCGCTGGCC----TGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGG 582
              || |||||| | |||     || ||||  |  | ||    |  || | | | || | | |
Db        303 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGATCGGTCGCCACGTG 362

Qy        583 GACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 642
              |   |  ||   |||   | || |||| |   | |   |||  ||||  |   | |   |
Db        363 GCGTGATCGGTCGCCACGTAGCGTGATCGGTCGCCACGTGGCGTGCTCGGTCGCCACGTA 422
```

```
Qy        643 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTG 702
              |  || | |   | |   | |  |||| |   | |   | |  |||| |   | | | |
Db        423 GCGTGTTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 482

Qy        703 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTG 762
              |  |||| |   | |   | |  |||| |   | | | | |  |||| |   | |   |
Db        483 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 542

Qy        763 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCT 822
              |  |||| |   | |   | |  |||| |   | |   | |  |||| |   | |   |
Db        543 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 602

Qy        823 GCTGGCTTGCTTGCTTGCTTGCTAGTGGG 851
              ||  ||| | | |||   | || |  | |
Db        603 GCGTGCTCGGTCGCTACGTAGCGACCGTG 631


RESULT 16
BH537125
LOCUS       BH537125                597 bp    DNA     linear   GSS 14-DEC-2001
DEFINITION  BOHCS81TR BOHC Brassica oleracea genomic clone BOHCS81, genomic
            survey sequence.
ACCESSION   BH537125
VERSION     BH537125.1  GI:17772878
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 597)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOHCS81TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .597
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHCS81"
                     /clone_lib="BOHC"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       34 a    180 c    237 g    146 t
ORIGIN

  Query Match             12.7%;  Score 127.2;  DB 28;  Length 597;
  Best Local Similarity   52.6%;  Pred. No. 9.4e-14;
  Matches  302;  Conservative    0;  Mismatches  268;  Indels    4;  Gaps    1;

Qy        223 GCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 282
              || |||  | ||||||    || ||||  |  |       || |||| | ||||||
Db          5 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 64

Qy        283 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 342
              || ||||  |  |       || |||| | |    |   ||   ||  | | |||   |
Db         65 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 124

Qy        343 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCT 402
              || |||| | |   |   ||   | | |||   |||   |   | ||||||
Db        125 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 184

Qy        403 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 462
              || |||| | ||||||    || |||| | ||||||    || |||| | ||||||
Db        185 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 244

Qy        463 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 522
              || |||| | ||||||    || |||| | ||||||    || |||| | ||||||
Db        245 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 304

Qy        523 GCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGG 582
              || |||| | ||||||    || |||| |  |       |  |||| | | |||  |
Db        305 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 364

Qy        583 GACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 642
              |   |  ||   ||   | || || |     ||  |||| |  ||  |   ||  ||||
Db        365 GCGTGCTCGGTCGCTGCGTAGCGTGCT----CGGTCGCTGCGTAGCGTGCTCGGTCGCTG 420

Qy        643 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTG 702
              |  ||  |   ||  |||| |  ||  |   ||  |||| |  ||  |   ||  |||
Db        421 CGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTG 480

Qy        703 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTG 762
              |  ||  |   ||  |||| |  ||  |   ||  ||| |  ||  |   ||  ||||
Db        481 CGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTG 540

Qy        763 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 796
              |  ||  |   ||  |||  |  ||  ||||||
Db        541 CGTAGCGTGCTCGGTCGCTACGTAGCGCGTAGGG 574


RESULT 17
CNS03LX7
LOCUS       CNS03LX7                743 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            036P11 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL250036
VERSION     AL250036.1  GI:7971048
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 743)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
```

```
FEATURES             Location/Qualifiers
     source          1. .743
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="036P11"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG036CH06SP1-end :
                     PUC-Ori"
BASE COUNT        0 a    191 c    360 g    192 t
ORIGIN
```

Query Match 12.7%; Score 127.2; DB 29; Length 743;
Best Local Similarity 60.1%; Pred. No. 9.4e-14;
Matches 441; Conservative 0; Mismatches 273; Indels 20; Gaps 13;

```
Qy        219 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 278
              ||||| ||| |||   ||||||  ||| |||| |  | | |    | ||||| ||| ||||
Db         14 GGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTG-GCTGGCTGGCTGGCTGCTGGCTGGCTG 72

Qy        279 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 338
              ||  ||| ||||   ||     |  | |||  ||     |  | | |   | ||||| |||
Db         73 GCTGGCTGGCTGGCTGGCTGCT--GGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTG 130

Qy        339 GCTGGCCTGCTTGCTGAGGGGACAGTAG--GGCTGCTGGCTTGCTAGTAGGGCTGCTCGC 396
              ||||||  ||| ||||   ||     | |  |||||   |  ||   |  ||||||| ||
Db        131 GCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGC 190

Qy        397 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 456
              ||||  ||| ||| ||| |||||  |||||  ||| ||| ||||||  ||| ||| ||| ||
Db        191 TGGCTGGCTGGCTGGCTGGCTGG-CTGCTGGCTGGCTGGCTGGCTGGCTGGCT-GC 248

Qy        457 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 516
              ||||  ||| ||| ||| |||||  |||  |||| ||| ||||||  ||| ||| ||| ||
Db        249 TGGCTGGCTGGCTGGCTGGCTGG-CTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGC 307

Qy        517 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGG--GACAGTAGGGCTGCTTGCTT 574
              | ||  ||| ||| ||| ||||||  ||| ||||  ||  | | |   |||||  ||  |
Db        308 T-GCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCT 366

Qy        575 GCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAG--GGCTGCTGGGCTGCTAGT 632
              |       ||  ||   ||  ||  | |||||  | |  | |  |||||   |||||   |
Db        367 GGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGG 426

Qy        633 AGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG--GGCTGCTGGGCTGCTA 690
                |||||||||   ||| |  ||   |||||   |||  | |  |||||   |||||
Db        427 CTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCT 486

Qy        691 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG--GGCTCCTGGGCTGC 748
              |   |||| ||||   ||| |  ||   |||||   |||  | |  ||||   |||||
Db        487 GGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGG 546

Qy        749 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 808
                |  |||||||||   ||| |  ||   |||||   |||  |  |||||   |||||
Db        547 CTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCT--GGCTGGCTGGCTGG 604

Qy        809 TAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACT 868
                |  ||||| |||||||||| ||| ||| ||| ||| |   ||| ||| ||  |||
Db        605 CTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCT-GGCTGGCTGCTGGCTGGCTGGC 663

Qy        869 AGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGGGCTGCCGGCCAGCTGCGTGCGAGGGCG 928
              ||   ||||    ||||   |       |  ||||| | | | |     ||   || |
Db        664 TGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTG 723

Qy        929 TCCGAGGGCCAGAC 942
              | |   | | | |
Db        724 GCTGGCTGGCTGGC 737
```

```
RESULT 18
DR37P5T/c
LOCUS       DR37P5T                  640 bp    DNA     linear   GSS 22-NOV-2002
DEFINITION  Danio rerio genomic clone DKEY-37P5, genomic survey sequence.
ACCESSION   AL975317
VERSION     AL975317.1  GI:25181115
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 640)
  AUTHORS   Humphray,S.J., Huckle,E. and Hunt,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2002) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 37P5. 37P5 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene.
            Further details: http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .640
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-37P5"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
BASE COUNT      299 a    232 c     55 g     54 t
ORIGIN
```

Query Match 12.4%; Score 124.4; DB 29; Length 640;
Best Local Similarity 54.7%; Pred. No. 3e-13;
Matches 289; Conservative 0; Mismatches 236; Indels 3; Gaps 2;

```
Qy        349 TTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGC 408
              |||||  |   |   ||     || | | | ||| |    | |||| | |  | || |||
Db        639 TTGCTTACCTGGTTGTTTCCTTGGTTGTTAGCTTGGTTAGTTGCTTGATTACTTGGTTGG 580

Qy        409 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 468
              ||| | | | |  || ||||||    |||    || ||  ||||| ||| || ||||||
Db        579 TTGTTTGATTGGTTGGTTGCTTTGATGCTTAGTTGGTTAGTTGCTTGCTTGCTTGCTTGA 520

Qy        469 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 528
              | | |  ||||   | ||| ||  | ||| || ||||||||||| | | || |  | | |
Db        519 TAGTTGACTGGTTAG-TTGGTTAGTTGCTTGCTTGCTTGCTTGATAGTTGACTGGTTAGT 461

Qy        529 TTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAG--GGCTGCTTGCTTGCTAAGGGGACG 586
              | | | |||||   | | | ||    |   || |  || |||||||||| || |  |
Db        460 TGGTTGGCTGGTTGGTTGGTTGGTTAGTTGGTTGTTGGTTGCTTGCTTGGTATGTTGGTT 401

Qy        587 GTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 646
                || ||  || | ||| | | |   || || | || || | |   || || |  | |
Db        400 AGGTGCTTGGTTGGTTGCTAGCTTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTAGTT 341

Qy        647 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCT 706
               || |   || || | |  || |     || ||    | || | ||  || |  | || |
Db        340 ACTTGGTTGGTTGGTTGATTGGTTTCTTGGTTGGATAGTTGGTTGTTTGGTTGGTTGGTT 281

Qy        707 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCT 766
              | |      |  || | || |||| |   || | ||  || | |   || || | || |
Db        280 GGTTAGTTGTTTGGTTGGTTGCTTGGTAGGTTGCTTACTTGGTTGGTTGGATGGTTGGTT 221

Qy        767 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTG 826
              | | |   || || | || || | |   || |  | || |||| |  ||| ||  || |
Db        220 GGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTTGTTGGTTGCTTGGTGGGTTGAGTGGTT 161

Qy        827 GCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCT 874
              | |||||| |||| ||| |      ||  | | |  || |     ||||
Db        160 GGTTGCTTACTTGGTTGGTTTCTTGGTTGGTCGGTTGGTTGGTTGGCT 113
```

```
RESULT 19
AZ828483/c
LOCUS       AZ828483                 481 bp    DNA     linear   GSS 20-FEB-2001
```

```
DEFINITION  2M0105A06R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0105A06 R, genomic survey sequence.
ACCESSION   AZ828483
VERSION     AZ828483.1  GI:12998391
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 481)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0105  row: A  column: 06
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 481.
FEATURES             Location/Qualifiers
     source          1. .481
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0105A06"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      148 a    137 c    116 g     80 t
ORIGIN

  Query Match             12.3%;  Score 123.4;  DB 28;  Length 481;
  Best Local Similarity   73.7%;  Pred. No. 4.6e-13;
  Matches  157;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy        390 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 449
              |||| ||| ||  ||||||| ||| |||||| |||| ||||||| ||| ||  ||||||| 
Db        450 TGCTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCT 391

Qy        450 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 509
                ||| |||||| |||| ||||||| ||| ||  ||||||| ||| |||||| |||| ||| 
Db        390 GGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCTGGCT 331

Qy        510 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 569
              |||| ||| ||  ||||||| ||| |||||| |||| |||    ||   |  ||  |||| 
Db        330 TGCTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCTGGCTTGCT 271

Qy        570 TGCTTGCTAAGGGGACGGTACGCCTGCCTGATG 602
                |||||||| ||   | |  | |    ||||||
Db        270 GGCTTGCTATGGACCCAGAGCTCTGTGCTGATG 238


RESULT 20
AK028099/c
LOCUS       AK028099               1142 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus adult male testis cDNA, RIKEN full-length enriched
            library, clone:1700125N06 product:unclassifiable, full insert
            sequence.
ACCESSION   AK028099
VERSION     AK028099.1  GI:26080728
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
```

```
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 1142)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .1142
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:1700125N06"
                     /db_xref="taxon:10090"
                     /clone="1700125N06"
                     /sex="male"
                     /tissue_type="testis"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     misc_feature    1. .1142
                     /note="unclassifiable"
BASE COUNT      303 a    340 c    271 g    228 t
ORIGIN

  Query Match           12.2%;  Score 122.4;  DB 11;  Length 1142;
  Best Local Similarity   72.3%;  Pred. No. 7.2e-13;
  Matches  159;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy        606 TGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGC 665
              || | |    |||||||||||||| |   |||||||||||||| |    | ||||| ||
Db        436 TGCTGGGCAAGCTGCTGGGCTGCTGGGCAGGCTGCTGGGCTGCTGGGCAAGTTGCTGAGC 377

Qy        666 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGC 725
              ||||  |   |||||||||||||| |   || | |||||||||| |   |||||||||||
Db        376 TGCTGGGCAGGCTGCTGGGCTGCTGGACAGGATGCTGGGCTGCTGGGCAGGCTGCTGGGC 317

Qy        726 TGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGC 785
              |||| | | || | |||||||||| |   || |||||| |||| |   || |||||||
Db        316 TGCTGGGAAGGTTGCTGGGCTGCTGGGCAGGGTGCTGGGATGCTGGGTAGGGTGCTGGGA 257

Qy        786 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCT 825
              |||| |   || ||||||| |||| |   || ||| || |
Db        256 TGCTGGGTAGGGTGCTGGGATGCTGGGTAGGGTGCTTGGT 217
```

```
RESULT 21
BH486116
LOCUS       BH486116                599 bp    DNA     linear   GSS 13-DEC-2001
DEFINITION  BOGJZ09TF BOGJ Brassica oleracea genomic clone BOGJZ09, genomic
            survey sequence.
ACCESSION   BH486116
VERSION     BH486116.1  GI:17694220
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 599)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOGJZ09TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .599
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGJZ09"
                     /clone_lib="BOGJ"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       63 a    178 c    208 g    150 t
ORIGIN

  Query Match           12.2%;  Score 122.2;  DB 28;  Length 599;
  Best Local Similarity   52.1%;  Pred. No. 7.6e-13;
  Matches  305;  Conservative    0;  Mismatches  268;  Indels   12;  Gaps    1;

Qy        226 TGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCT 285
              ||   | |||||     || ||||  |  |  |    || |||| | |||||    .||
Db         19 TGTTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCG 78

Qy        286 TGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 345
              ||||  |  |  |    || |||| | |         ||   ||  | | |||   | ||
Db         79 TGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCG 138

Qy        346 TGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCT 405
              |||| | |         ||   |   | | |||   |||   |   | |||||    ||
Db        139 TGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCG 198

Qy        406 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 465
              |||| | |||||     || |||| | |||||     || |||| | |||||    ||
Db        199 TGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCG 258

Qy        466 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 525
              |||| | |||||     || |||| | |||||     || |||| | |||||    ||
Db        259 TGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCG 318

Qy        526 TGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGAC 585
              |||| | |||||     || |||| |  | |  |     |  |||| | | |||| |
Db        319 TGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACG----- 373

Qy        586 GGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGC 645
                    || || | | | ||    |  |||| ||  ||||    |  |||| ||
Db        374 -------TAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGT 426
```

```
Qy        646 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGC 705
              ||||    |  |||| ||  ||||    |  |||| ||  ||||    |  | || ||
Db        427 CGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGT 486

Qy        706 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGC 765
              ||||    |  |||| ||  ||||    |  | || ||  ||||    |  |||| ||
Db        487 CGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGA 546

Qy        766 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 810
              ||||    |  |||| ||  ||||    |  |||| ||  ||||
Db        547 CGCTACGTAGCGTGCTCGGACGCTACGTAGCGTGCTCGGTCGCTA 591


RESULT 22
AL528199
LOCUS       AL528199                1100 bp    mRNA    linear   EST 23-MAY-2003
DEFINITION  AL528199 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DC025YF13 5-PRIME, mRNA sequence.
ACCESSION   AL528199
VERSION     AL528199.2  GI:31066049
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1100)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 13, 2001 this sequence version replaced gi:12791692.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www:genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 1419.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DC025CC07QP1&cluster=1419.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DC025CC07QP1.
FEATURES             Location/Qualifiers
     source          1. .1100
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DC025YF13"
                     /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT       74 a    296 c    463 g    258 t      9 others
ORIGIN

  Query Match           12.2%;  Score 122;  DB 9;  Length 1100;
  Best Local Similarity   61.7%;  Pred. No. 8.5e-13;
  Matches  400;  Conservative    0;  Mismatches  225;  Indels   23;  Gaps   12;

Qy        224 CCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTG 283
              | |||  ||| ||||||  ||| ||||  |  | | |  |||  ||| ||| ||||||  |
Db        245 CTTGCTGGCTGGCTGGCTGGCTG-GCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGG 303

Qy        284 CTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGG 343
              || ||||  ||  | |   | |||  ||    |  | | |  |||  ||| ||| |||||
Db        304 CTGGCTGCTGG--CTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGG 361

Qy        344 CCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTG 403
              |  ||| ||||  ||        |||||   |  ||   |  ||||||| ||||||  |
Db        362 CTGGCTGGCTGCTGG------CTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGG 415

Qy        404 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 463
              || ||| ||| ||||| ||||| ||| ||| |||||||  ||| ||| ||| |||||||  |
Db        416 CTGGCTGGCTGGCTGG-CTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT-GCTGGCTGG 473

Qy        464 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 523
              || ||| ||| ||||| |||  |||| ||| |||||||  ||| ||| ||| ||| ||  |
Db        474 CTGGCTGGCTGGCTGG-CTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT-GCTGG 531

Qy        524 CTTGCTTGCTCGCTGGCCTGCTTGCTGAGGG--GACAGTAGGGCTGCTTGCTTGCTAAGG 581
              || ||| ||| |||||||  ||| ||||  ||  | | |   |||||  ||  ||
Db        532 CTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGCT 591

Qy        582 GGACGGTACGCCTGCCTGATGGCTTGATAGTAG--GGCTGCTGGGCTGCTAGTAGGGCTG 639
              ||  ||   ||  ||  | ||||| | |  | |  |||||   |||||   |   |||||
Db        592 GGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 651

Qy        640 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG--GGCTGCTGGGCTGCTAGTAGGGC 697
              ||||   ||| |  ||   |||||   |||  | |  |||||   |||||   |   |||
Db        652 CTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGC 711

Qy        698 TCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG--GGCTCCTGGGCTGCTAGTAGG 755
              | ||||   ||| |  ||   |||||   |||  | |  ||||    |||||   |   |
Db        712 TGCTGGCTGGCTGGCTGGCTGGCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGG 771

Qy        756 GCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGG 815
              ||||||||   ||| |  ||   |||||   |||  |  |||||   |||||   |   |
Db        772 GCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGCT--GGCTGGCTGGCTGGCTGGCTG 829

Qy        816 GCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTG 863
              ||||  |||||||| ||| ||| ||| ||| |  ||     | || ||
Db        830 GCTGGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGCATGGCATG 877


RESULT 23
BZ041698
LOCUS       BZ041698                729 bp    DNA     linear   GSS 09-OCT-2002
DEFINITION  llr23h12.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
            sequence.
ACCESSION   BZ041698
VERSION     BZ041698.1  GI:23631988
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 729)
  AUTHORS   Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash
            ,W., Rabinowicz,P.D. and Wilson,R.K.
  TITLE     Whole genome shotgun reads from Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: llr23  row: h  column: 12
            Seq primer: -21UPpOT forward
            Class: shotgun
            High quality sequence start: 8
            High quality sequence stop: 551.
FEATURES             Location/Qualifiers
     source          1. .729
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3712"
                     /clone_lib="B.oleracea002"
                     /note="Vector: pOTw13; Whole genome shotgun library from
                     flowering buds. DNA was purified  from a crude nuclear
                     prep using Brassica oleracea TO1000DH3 buds  provided by
                     Thosmas Osborn at the University of Wisconsin. Genomic
                     DNA was provided by Pablo Rabinowicz (CSHL) and the
                     shotgun library  prepared at Washington University Genome
                     Sequencing Center."
```

```
BASE COUNT      114 a    237 c    235 g    141 t      2 others
ORIGIN

  Query Match            12.0%;  Score 120.4;  DB 28;  Length 729;
  Best Local Similarity   53.1%;  Pred. No. 1.6e-12;
  Matches  279;  Conservative    0;  Mismatches  242;  Indels    4;  Gaps    1;

Qy        220 GCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGG 279
              || || |||  | |||||   | || ||||  |  |  |    || |||||| |||||
Db        207 GCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTTGGTCGCTAC 266

Qy        280 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCG 339
              | || |||| |  |  |    || |||| | |      |||   ||  | | |||
Db        267 GCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTAC 326

Qy        340 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGG 399
                || |||| | |       |||   | | | | |||    ||   | | | |||||
Db        327 GCAGCGTGCTCGGTCGCTACGCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTAC 386

Qy        400 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 459
              | || |||  | |||||   | || |||  | |||||   | || |||  | |||||
Db        387 GCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTAC 446

Qy        460 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 519
              | || |||  | |||||   | || |||  | |||||   | || |||  | |||||
Db        447 GCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTAT 506

Qy        520 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAA 579
              | || |||  | |||||   | || |||   |  |   |   |  |||  | | ||||
Db        507 GCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTACGCAGCATGCCCGGTCGCTAC 566

Qy        580 GGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 639
              |   | | |  || |||  | | |||     |       |||| ||  ||||    |  ||
Db        567 GCACATGCTCAGCGTGCTCGGTCGCTACGCANCG----TGCTCGGTCGCTACGCAGCGTG 622

Qy        640 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTC 699
              || ||  ||||    |  |||| ||  ||||    |  |||| ||  ||||    |  |
Db        623 CTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGAGTG 682

Qy        700 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGG 744
              || ||  ||||    |  || | |  ||||  | |  | || ||
Db        683 CTCGGTCGCTACGCAGCGTGGTCCGGCGCTACGACGCGTGCTCGG 727


RESULT 24
AZ650539
LOCUS       AZ650539                693 bp    DNA     linear   GSS 14-DEC-2000
DEFINITION  1M0520I14R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0520I14 R, genomic survey sequence.
ACCESSION   AZ650539
VERSION     AZ650539.1  GI:11785128
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 693)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0520  row: I  column: 14
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 693.
FEATURES             Location/Qualifiers
     source          1. .693
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0520I14"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      144 a    156 c    166 g    226 t      1 others
ORIGIN

  Query Match            12.0%;  Score 120.2;  DB 28;  Length 693;
  Best Local Similarity   72.8%;  Pred. No. 1.8e-12;
  Matches  155;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        325 AGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGT 384
              |||| |||||||| ||| ||||||||||      | | |   | ||||  ||||||  |
Db        480 AGGCTTGCTTGCTTGCTTGCCTGCTTGCCTGCTTGCCTGCCTGCCTGCCTGCTTGCCTGC 539

Qy        385 AGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 444
                | ||||| ||  || ||| |||||||| ||| || ||||||||||| ||| || |||
Db        540 TTGTCTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 599

Qy        445 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 504
              | |||||||  ||  ||||||  ||| |||  ||  ||||||  ||| |||  |   || ||
Db        600 TAGCTTGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGACTGCTTGT 659

Qy        505 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 537
              |||||||  ||| ||||||||||||||| |||
Db        660 CTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCT 692


RESULT 25
AZ633613/c
LOCUS       AZ633613                591 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  1M0488O23R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0488O23 R, genomic survey sequence.
ACCESSION   AZ633613
VERSION     AZ633613.1  GI:11755803
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
```

```
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0488  row: O  column: 23
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 591.
FEATURES             Location/Qualifiers
     source          1. .591
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0488O23"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      243 a    201 c     61 g     86 t
ORIGIN

  Query Match             12.0%;  Score 119.8;  DB 28;  Length 591;
  Best Local Similarity   52.5%;  Pred. No. 2.1e-12;
  Matches  262;  Conservative    0;  Mismatches  237;  Indels    0;  Gaps    0;

Qy        367 GGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 426
              || ||   ||||||| ||  || ||||  |  || ||||||||||||| ||| || ||||||
Db        501 GGTTGGGTGCTTGCTTGTTTGGTTGCTTACATGCTTGCTTGCTTGCTTGCTTGCTTGCTT 442

Qy        427 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 486
              ||||||| ||| || |||||   || |   | | ||  ||| ||  | | |    || |
Db        441 GCTTGCTTGCTTGCTTGCTTAGATGATTAGTTGGCTCATTGGTTTGTTGGTTCATTGGTC 382

Qy        487 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 546
                | | | | | |   | ||| |   ||  | |  || ||| || | | |    || ||
Db        381 AGTCGGTTGGTCGGTCGGTTGGTCAGTCATTTGATTGGTTGGATGGTTGGTTAGTTGGTT 322

Qy        547 GCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTT 606
                | |  ||  |   |  ||  || |||       ||| |   |  ||  || | | ||
Db        321 AGTTGGCTGAGTGATTGCTTGGGTGATTGACTGAAAGACTGGTTGTTTGGTTGCTTGGTT 262

Qy        607 GATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 666
              |||  |   || ||   | ||| | |   || ||   ||||| | |   || || | || |
Db        261 GATTGCTTGGTTGGGTGACTGGTTGGTTGGGTGACTGGCTGGTTGGTTGGTTGTTTGGTT 202

Qy        667 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCT 726
              | | ||  || || | || || | |   || |  | || ||   |   |  |    | |
Db        201 GGTTGTTTGGTTGGTTGGTTGGTTGGTTGGTTGGGTGAGTGAGTGAGTCAGTGAGT 142

Qy        727 GCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 786
              |   |   |  |    |  || |         || | ||    | ||   | ||    | |
Db        141 GAGTGAGTGAGTGACTGCTTGGTTAGGCAATTGGTTGGTAAATTGTTTAGTTGGGTTGTT 82

Qy        787 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTA 846
              | | | |  | || |  |  | |   ||  |  || | ||||| |||||||  ||  |
Db         81 GGTTGAATGCTTCCTTGAGTAATTGGCTGGAAGATTGGTTGCTTGATTGCTTGGTTAGTT 22

Qy        847 GTGGGGCCGCTTGCCTGCT 865
              |   ||  |||||| ||||
Db         21 GGTTGGTTGCTTGCTTGCT 3


RESULT 26
BB610246/c
LOCUS       BB610246                659 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB610246 RIKEN full-length enriched, adult male testis Mus musculus
            cDNA clone 1700125N06 5', mRNA sequence.
ACCESSION   BB610246
VERSION     BB610246.1  GI:16451804
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 659)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
```

```
FEATURES             Location/Qualifiers
     source          1. .659
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="1700125N06"
                     /sex="male"
                     /tissue_type="testis"
                     /dev_stage="adult"
                     /lab_host="SOLR"
                     /clone_lib="RIKEN full-length enriched, adult male testis"
                     /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGAGCGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites. "
BASE COUNT      158 a    229 c    146 g    125 t      1 others
ORIGIN

  Query Match             12.0%;  Score 119.8;  DB 10;  Length 659;
  Best Local Similarity   71.4%;  Pred. No. 2.1e-12;
  Matches  157;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        606 TGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGC 665
              || | |    ||||||||| |||| |   ||||||||||||||| |    | ||||| ||
Db        433 TGCTGGGCAAGCTGCTGGGNTGCTGGGCAGGCTGCTGGGCTGCTGGGCAAGTTGCTGAGC 374

Qy        666 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGC 725
              |||| |   |||||||||||||||| |   || | |||||||||| |   ||||||||||
Db        373 TGCTGGGCAGGCTGCTGGGCTGCTGGACAGGATGCTGGGCTGCTGGGCAGGCTGCTGGGC 314

Qy        726 TGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGC 785
              |||| | | || | ||||| |||| |   || |||||| |||| |   || |||||||
Db        313 TGCTGGGAAGGTTGCTGGGGTGCTGGGCAGGGTGCTGGGATGCTGGGTAGGGTGCTGGGA 254

Qy        786 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCT 825
              |||| |   || ||||||| |||| |   || ||| || |
Db        253 TGCTGGGTAGGGTGCTGGGATGCTGGGTAGGGTGCTTGGT 214


RESULT 27
BH284699
LOCUS       BH284699               729 bp    DNA     linear   GSS 30-NOV-2001
DEFINITION  CH230-49O13.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
            CH230-49O13, genomic survey sequence.
ACCESSION   BH284699
VERSION     BH284699.1  GI:17197105
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 729)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 EcoRI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-49O13.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 49  row: O  column: 13
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .729
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-49O13"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 1"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      180 a    187 c    150 g    212 t
ORIGIN

  Query Match             12.0%;  Score 119.8;  DB 28;  Length 729;
  Best Local Similarity   79.3%;  Pred. No. 2.1e-12;
  Matches  142;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy        371 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 430
              ||| || |||| || ||  |||| | | || ||||||||||||| ||| || |||||||||
Db        510 GCTAGCCTGCTGGTGGGCTTGCTAGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 569

Qy        431 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              ||| ||| || ||||| |||||| ||| || |||||||||||| ||| || |||||||||
Db        570 GCTTGCTTGCTTGCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 629

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              ||| |||  | |||||||||||| ||| || ||||||| || | ||| || ||||  ||
Db        630 GCTTGCTTCCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGATGGCTAGCTTGCTATCT 688


RESULT 28
AZ782118
LOCUS       AZ782118               420 bp    DNA     linear   GSS 16-FEB-2001
DEFINITION  2M0022P01F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0022P01 F, genomic survey sequence.
ACCESSION   AZ782118
VERSION     AZ782118.1  GI:12915490
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 420)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
```

```
           Email: ddunn@genetics.utah.edu
           Insert Length: 10000   Std Error: 0.00
           Plate: 0022  row: P  column: 01
           Seq primer: CGTTGTAAAACGACGGCCAGT
           Class: plasmid ends
           High quality sequence stop: 420.
FEATURES             Location/Qualifiers
     source          1. .420
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0022P01"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       13 a    219 c     79 g    109 t
ORIGIN

  Query Match             11.8%;  Score 118.6;  DB 28;  Length 420;
  Best Local Similarity   57.0%;  Pred. No. 3.4e-12;
  Matches  217;  Conservative    0;  Mismatches  164;  Indels    0;  Gaps    0;

Qy        219 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 278
              | || ||||||| |  ||  |||||| |||      | | |   ||||||  |||  ||
Db         10 GCCTCCCTGCCTCCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCT 69

Qy        279 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 338
              ||||||  | |      | | |   |||| | |||      | | | |  ||||| |||
Db         70 GCCTGCCTCCCTCCCTGCCTGCCTGCCTACCTGCCTGCCTGCCTGCATCCCTGCCTGCCT 129

Qy        339 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 398
              ||   ||||| |||    | |     | ||||  || |||  |   | || |  ||
Db        130 GCCTACCTGCCTGCCTGCCTGCCTCCCTGCCTGCCTGCCTGCCTGCCTGCCTACCTGCCT 189

Qy        399 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 458
               ||||| | | |||   |  |||||| ||| |||  ||  |||||| | | |||   |
Db        190 CCCTGCCTCCCTGCCTACCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCTGCCTCCCT 249

Qy        459 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 518
              |||| | ||| | |  ||  |||||| ||| |||  ||  |||||| ||| | ||  |
Db        250 GCCTCCCTGCCTCCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTTCTATCCT 309

Qy        519 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTA 578
              |||| | | | | |  ||  |||||| |||      | |     | |||| ||| |||
Db        310 GCCTCCCTCCCTCCCTGCCTGCCTGCCTGCCTGCCTGCCTACCTGCCTGCCTGCCTGCCT 369

Qy        579 AGGGGACGGTACGCCTGCCTG 599
                  | | |   |||||||||
Db        370 GCCTGCCTGCCTGCCTGCCTG 390


RESULT 29
BZ285588
```

```
LOCUS       BZ285588                549 bp    DNA     linear   GSS 15-OCT-2002
DEFINITION  CH230-385O23.TJC CHORI-230 Segment 2 Rattus norvegicus genomic
            clone CH230-385O23, genomic survey sequence.
ACCESSION   BZ285588
VERSION     BZ285588.1  GI:24016713
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-385O23.TVB
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 385  row: O  column: 23
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-385O23"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      149 a    104 c    211 g     85 t
ORIGIN

  Query Match             11.7%;  Score 117.2;  DB 29;  Length 549;
  Best Local Similarity   53.7%;  Pred. No. 6.2e-12;
  Matches  288;  Conservative    0;  Mismatches  243;  Indels    5;  Gaps    2;

Qy        299 GTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCT-TGCTCGCTGGCCTGCTTGCTGAGG 357
              | | ||  || || | || ||    | |||||||  |   |    | ||    ||| | | ||
Db         14 GGATGCAAGCCTGATTAGAGGGGTGCAGGCCTGATCAGAGGGATGCAAGGCTGGATCAGA 73

Qy        358 GGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCT 417
              ||   | | | |||        |   ||| | ||| ||   ||  |||  || || ||
Db         74 GGGATGCAAGCCTGACCAGAGGGGTGTAAGCCTGATCAGAGGGGTGCAAGCCTGATCAGA 133

Qy        418 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 477
              ||  |||  |  || ||   ||  |||  || || ||   ||  |||  || || ||
Db        134 GGGATGCAAGGCTGATCAGAGGGATGCAAGCCTGATCAGAGGGGTGCAAGCCTGATCAGA 193

Qy        478 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC--- 534
              ||  |||  || || ||   ||  |||  || || ||   ||  |||  |  || ||
Db        194 GGGGTGCAAGCCTGATCAGAGGGATGCAAGCCTGATCAGAGGGATGCAAGGCTGATCAGA 253

Qy        535 -GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCC 593
               |   ||  || || | || ||   | | | |||  |     |       | | |  |
Db        254 GGGATGCAAGCCTGATCAGAGGGATGCAAGCCTGATCAGAGGGGTGCAAGCCTGATCAGA 313
```

```
Qy        594 TGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 653
              |  ||   || ||||   ||||  |||  | ||| |   ||||  ||   | |||  |   |
Db        314 GGGGTGCAAGCCTGATCAGAGGGGTGCAAGCCTGATCAGAGGGATGAAAGCCTGATCAGA 373

Qy        654 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTA 713
              |||  |||  | |||  |   ||||  |||  |||||  |   ||||  | |  |||||  |   |
Db        374 GGGGTGCAAGCCTGATCAGAGGGATGCAAGGCTGATCAGAGGGGTGCAAGGCTGATCAGA 433

Qy        714 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 773
              |||  |||  | |||  |   ||||  |  |  | |||  |   ||||  |||  |  |||  |   |
Db        434 GGGATGCAAGCCTGATCAGAGGGGTGCAAGCCTGATCAGAGGGGTGCAAGCCTGATCAGA 493

Qy        774 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCT 829
              |||  |||  | |||  |   ||||  ||   | |||  |   ||||  |||  ||  | |
Db        494 GGGGTGCAAGCCTGATCAGAGGGATGAAAGCCTGATCAGAGGGGTGCAAGCCTGAT 549
```

RESULT 30
BM950444

```
LOCUS       BM950444                711 bp    mRNA    linear   EST 14-MAR-2002
DEFINITION  UI-M-EH0p-buu-i-05-0-UI.r1 NIH_BMAP_EH0p Mus musculus cDNA clone
            IMAGE:5687116 5', mRNA sequence.
ACCESSION   BM950444
VERSION     BM950444.1  GI:19434034
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 711)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin, Univeristy of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .711
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5687116"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 18.5 dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_EH0p"
                     /note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured mRNa was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site. Double stranded
                     cDNA was size selected according to mRNA size fraction,
                     ligated with EcoR I adaptor, digested with Not I, and then
                     cloned directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     , is CAGCCACGAC. This library was created for the
                     University of Iowa Mouse Brain Molecular Anatomy Project
                     (BMAP): 'Gene Discovery in the Developing Mouse Nervous
                     System', supported by National Institute of Mental Health
                     (NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT      177 a    155 c    271 g    108 t
```

ORIGIN

```
  Query Match           11.6%;  Score 116.2;  DB 12;  Length 711;
  Best Local Similarity  52.3%;  Pred. No. 9.5e-12;
  Matches  282;  Conservative    0;  Mismatches  253;  Indels    4;  Gaps    1;

Qy        286 TGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 345
              |||| | |   ||  | ||  |     |   | |  | | |||   ||  | |      |
Db         43 TGCTTATGTAGCACCATGCAGGGAAACCAGGTGTGCCGCAGGGAAGCCAGGTGTGCTTCA 102

Qy        346 TGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCT 405
              |   ||   | | || | ||||  ||  | |     | ||||  ||  | ||  ||||
Db        103 GGGAAGCCAGGTGTACTGCAGGGAAGCCAGGTGTACTGCAGGGAAGCCAGGTGTGCTGCA 162

Qy        406 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 465
              |   ||  | ||  ||||  |   ||  | ||  ||||  |   ||  | ||  | ||
Db        163 GGGAAGCCAGGTGTACTGCAGGGAAGCCAGGTGTACTGCAGGGAAGCCAGGTGTGCAGCA 222

Qy        466 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 525
              |   ||  | ||  || |  |   ||  | ||  ||||  |   ||  | ||  ||||
Db        223 GGGAAGCCAGGTGTGCTTCAGGGAAGCCAGGTGTACTGCAGGGAAGCCAGGTGTGCTGCA 282

Qy        526 TGCTTGCTCGCTGGCCTGCTT----GCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGG 581
              |   ||  | ||  ||||      ||   | | || | ||||  ||  | |    |
Db        283 GGGAAGCCAGTTGTGCTGCAGGGAAGCCAGGTGTACTGCAGGGAAGCCAGGTGTGCAGCA 342

Qy        582 GGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCT 641
              ||   |   |     |||  ||   |  ||  | ||||| |||   | || || |||||
Db        343 GGGAAGCCAGGTGTGCTGCAGGGAAGCCAGGTGTGCTGCAGGGAAACCAGGAGTGCTGCA 402

Qy        642 GGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCT 701
              |||  ||  ||  | |||||  |||   | || ||  |||||  ||   ||  ||  | ||| |
Db        403 GGGAAGCCAGGTGTGCTGCAGGGAAACCAGGAGCGCTGCAGGTAAGCCAGGTGCGCTGCA 462

Qy        702 GGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCT 761
              |||  ||  ||  | |||||  |||  ||  ||  | |||  | |||  ||  ||  | |||||
Db        463 GGGAAGCCAGGTGTGCTGCAGGGAAGCCAGGTGTGCTGCAGGGAAGCCAGGTGTGCTGCA 522

Qy        762 GGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 820
              ||   ||  ||  | |||||  |||  ||  ||  | |||||  ||   ||  ||  | |||||
Db        523 GGAAAGCCAGGTGCGCTGCAGGGAAGCCAGGTGTGCTGCAGGAAAGCCAGGTGTGCTGC 581
```

RESULT 31
BX426381

```
LOCUS       BX426381                1201 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX426381 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
            clone CS0DG006YM04 5-PRIME, mRNA sequence.
ACCESSION   BX426381
VERSION     BX426381.1  GI:30784544
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0AG006BG02QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DG006YM04"
```

```
                     /tissue_type="B CELLS (RAMOS CELL LINE)"
                     /cell_line="RAMOS CELL LINE"
                     /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
BASE COUNT        7 a    571 c    208 g    215 t    200 others
ORIGIN

  Query Match             11.5%;  Score 114.8;  DB 13;  Length 1201;
  Best Local Similarity   38.8%;  Pred. No. 1.7e-11;
  Matches  212;  Conservative   86;  Mismatches  248;  Indels    0;  Gaps    0;

Qy          4 CGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTC 63
              ||| ||   : |||  |   |||  ||   | |:| | | |||   | ||   |   |
Db        130 CGCGCGCKCKCGCGCGCGCGCGCGCGCGCGCKCGCGCGCGCGCGCGCGCGCGCGCGCG 189

Qy         64 TCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGC 123
                ||||   || | | | |  || |  |  | ||||    ||| |  |  ||||| |
Db        190 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 249

Qy        124 GCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGG 183
                |   |  | || | |  || | |  || |    |||| | |     | |  |  |
Db        250 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 309

Qy        184 GGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTG 243
                |  ||| | ||| ||  | | |    || |  ||   ||  ||  || ||| ||  |
Db        310 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 369

Qy        244 CTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGG 303
                |  ||    | |   |   ||  ||  || |||  ||  ||  ||    | |   |  |
Db        370 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 429

Qy        304 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 363
                |  ||  ||    | :   |   :|  :|: ||:|:|: :|  :|: :|:   : :   :
Db        430 CGCGCGCGCGCGCGCKCGCGCGCKCKCKCKCGCKCKCKCKCKCKCKCKCKCKCKCKCK 489

Qy        364 TAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 423
                :   :|: ||  ||: |   |   :|:|:|: :|  :|: :|: :|:|:|: :|  :
Db        490 CKCKCKCKCKCGCGCGCKCGCGCGCGCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCK 549

Qy        424 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 483
                |: :|: :|:|:|: :|  :|: :|: :|:|:|: :|  :|: :|: :|:|:|: ||  :
Db        550 CKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCGCKCK 609

Qy        484 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 543
                |: :|: :|:|:|: :|  :|: :|: :|:|:||  :|  :.|: :|: :|:| ||  |
Db        610 CKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCTCKCKCKCKCKCKCKCTCTCTCKCT 669

Qy        544 CTTGCT 549
                || :|:
Db        670 CTCKCK 675


RESULT 32
BZ015565
LOCUS       BZ015565                726 bp    DNA     linear   GSS 07-OCT-2002
DEFINITION  oeh87c12.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
            sequence.
ACCESSION   BZ015565
VERSION     BZ015565.1  GI:23572725
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 726)
  AUTHORS   Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash
            ,W., Rabinowicz,P.D. and Wilson,R.K.
  TITLE     Whole genome shotgun reads from Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: oeh87  row: c  column: 12
            Seq primer: -21UPpOT forward
            Class: shotgun
            High quality sequence start: 45
            High quality sequence stop: 398.
FEATURES             Location/Qualifiers
     source          1. .726
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3712"
                     /clone_lib="B.oleracea002"
                     /note="Vector: pOTwl3; Whole genome shotgun library from
                     flowering buds. DNA was purified  from a crude nuclear
                     prep using Brassica oleracea TO1000DH3 buds  provided by
                     Thosmas Osborn at the University of Wisconsin. Genomic
                     DNA was provided by Pablo Rabinowicz (CSHL) and the
                     shotgun library  prepared at Washington University Genome
                     Sequencing Center."
BASE COUNT       51 a    250 c    271 g    153 t      1 others
ORIGIN

  Query Match             11.4%;  Score 114.6;  DB 28;  Length 726;
  Best Local Similarity   53.1%;  Pred. No. 1.9e-11;
  Matches  269;  Conservative    0;  Mismatches  234;  Indels    4;  Gaps    1;

Qy        378 TGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 437
              | ||||   |   |  |||||    || |||| |  |||||    || |||| |  ||||
Db         35 TTCTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCCGCT 94

Qy        438 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 497
              |    || |||| |  |||||    || |||| |  |||||    || |||| |  ||||
Db         95 GCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCCGCT 154

Qy        498 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGAC 557
              |    || |||| |  | |||    || |||| |  |||||    || |||| |  |
Db        155 GCGTAGCGTGCTCGGCCACTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCCGCT 214

Qy        558 AGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGC 617
                  |  |||| |   |||  |  |   |  || | ||    | || || |     |||
Db        215 GCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCT----CGGC 270

Qy        618 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 677
              ||||  |  ||  |  |||  ||||  |  ||  |  |||  ||||  |  ||  |  |||
Db        271 AGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGC 330

Qy        678 TGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 737
              ||||  |  ||  |  |||  |||  |  ||  |  |||  ||||  |  ||  |  |||
Db        331 CGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGC 390

Qy        738 TCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 797
                |||  |  ||  |  |||  ||||  |  ||  |  |||  ||||  |  ||  |  |||
Db        391 CGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGC 450

Qy        798 TGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 857
              ||||  |  ||  |  |||  ||    | ||  |||| |   |||   |||  |  |  ||
Db        451 CGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGCCGCTGCGTAGCGTGCTCGGC 510

Qy        858 TGCCTGCTACTAGGGCTGCTGTGCAGC 884
                ||    ||   | |||     || ||
Db        511 CGCTGCGTAGCCGTGCTCGGCCGCTGC 537


RESULT 33
CA048466
LOCUS       CA048466                669 bp    mRNA    linear   EST 04-MAR-2003
DEFINITION  ssalsrkc009065 kidney Salmo salar cDNA, mRNA sequence.
ACCESSION   CA048466
VERSION     CA048466.1  GI:24354636
```

```
KEYWORDS    EST.
SOURCE      Salmo salar (Atlantic salmon)
  ORGANISM  Salmo salar
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   GRASP Consortium, Davidson,W.S., Koop,B.F. and
            http://web.uvic.ca/cbr/grasp.
  TITLE     A survey of Salmo salar transcripts from high complexity cDNA
            libraries
  JOURNAL   Unpublished
COMMENT     Contact: Koop BF
            Centre for Biomedical Research
            University of Victoria
            PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
            Tel: 250 472 4067
            Fax: 250 472 4075
            Email: bkoop@uvic.ca
            Centre for Biomedical Research, University of Victoria
            cDNA preparation and sequencing:
            Roberto Alberto, Marianne Beetz-Sargent, Maura Busby,
            Peter Hunt, Linda McKinnel, BF Koop.
            bioinformatics:
            Gordon D Brown.
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Salmo salar"
                     /mol_type="mRNA"
                     /strain="McConnell"
                     /db_xref="taxon:8030"
                     /clone_lib="kidney"
                     /note="Vector: pBlueScriptIISK+; Library Creator:  Matthew
                     L Rise ; Atlantic salmon tissue contributors: Carlo Biagi,
                     Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.), Simon
                     Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery (Crofton,
                     B.C.), Rachel Roper (University of Victoria)"
BASE COUNT      112 a    114 c    237 g    206 t
ORIGIN

  Query Match             11.4%;  Score 114.4;  DB 14;  Length 669;
  Best Local Similarity   50.5%;  Pred. No. 2e-11;
  Matches  303;  Conservative    0;  Mismatches  296;  Indels    1;  Gaps    1;

Qy        267 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 326
              ||| |||   ||||  ||| | |||  ||  |   | |  ||| ||||          |
Db         10 GCTGGCTGAATGGCAGGCTGGATGACTGGTTAACTGACTGGCTGGCTGGTTAACTGACTG 69

Qy        327 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 386
              ||  |||   | || ||| | || |||||  |||     ||    |||| |  ||    |
Db         70 GCTGGCTGAATGGCAGGCTTACTGGCTGACTGGATGCCTGGTTGACTGGTTAACTGACTG 129

Qy        387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 446
              |   ||| | || |  | |  ||  || ||||||  | | ||| | |  ||| |  |||
Db        130 GCTGGCTGGTTGACTGGTTAACTGACTGGCTGGCTGGTTGGCTGGATAACTGACTCGCTG 189

Qy        447 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 506
              ||| | |  ||||    ||   | | | |||||  | |  || | |  ||| |  | |
Db        190 GCTGGTTGACTGGTTAACTGAGTGGATGGCTGGTTGGTTGACTAGTTAACTGACTGGATG 249

Qy        507 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 566
              ||| | |  ||||    ||  || | | ||| |  ||  || | |   ||  | | ||||
Db        250 GCTGGTTGACTGGTTAACTGACTGGATGGCTTG-ATGACTGGTTAACTGAGTGGATGGCT 308

Qy        567 GCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCT 626
              | ||| |||    |    | |   | ||| ||| | |  || |     | |||   ||||
Db        309 GGTTGGTTGACTGGTTACCTGACTGGCTGGCTGGTTGACTGGTTAACTGACTGGATGGCT 368

Qy        627 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 686
              | | |   || |    | |||   |   || ||   || |    |   |||||   || |
Db        369 GGTTGACTGGTTAACTGACTGGATGGCTGGATGACTGGTTCACTGACTGGCTGGCTGGTT 428

Qy        687 GCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCT 746
              |  |     ||    |||||   |     |||   |||||   | | || |    || |
Db        429 GACTGGTTCACTGACTGGCTGGCTGGTTAACTGACTGGCTGGCTGAATGGATGGCTGGTT 488

Qy        747 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 806
              |   |     |||   |||||   |   | |||  |   |||   | | |||||  | | ||
Db        489 GACTGGTTAACTGACTGGCTGGCTGGTTGACTGGTTAACTGACTGGATGGCTGGTTGACT 548

Qy        807 GCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTA 866
              | |      | |||  ||   | |||  || ||   ||   |   ||| |  || ||| ||
Db        549 GGTTAACTGACTGGATGGCTGGTTGACTGGTTAACTGACTGGATGGCTGGATGACTGGTA 608


RESULT 34
AZ608625/c
LOCUS       AZ608625                680 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  1M0433A01F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0433A01 F, genomic survey sequence.
ACCESSION   AZ608625
VERSION     AZ608625.1  GI:11730815
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 680)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0433  row: A  column: 01
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 680.
FEATURES             Location/Qualifiers
     source          1. .680
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0433A01"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
```

```
                    and selected for ampicillin resistance."
BASE COUNT      253 a    223 c    113 g     91 t
ORIGIN

  Query Match             11.4%;  Score 114.2;  DB 28;  Length 680;
  Best Local Similarity   58.5%;  Pred. No. 2.2e-11;
  Matches  217;  Conservative    0;  Mismatches  153;  Indels    1;  Gaps    1;

Qy        349 TTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGC 408
              |||||||   |    | ||| |   |  | | | |   |||||   | |||  || |||
Db        498 TTGCTGATCAGTGGTTTGGGTTTGAGTTTGGGTTGGTTGGCTGGCTGGTGGGTTGGTTGG 439

Qy        409 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 468
               ||||   | || || | | | | | ||||| ||| ||  ||||| ||| || |||||||
Db        438 CTGCTTAGTTGCTTGTTGGTTGGTTGGCTGG-CTGGTTAGTTGCTTGCTTGCTTGCTTGC 380

Qy        469 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 528
              ||||| ||| || ||||||||||||| ||| || ||||||||||||| ||| || ||
Db        379 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGGCAAG 320

Qy        529 TTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGT 588
              ||| | | | | |||  || |  |  | | |   ||||| ||| ||| |  |  | | |
Db        319 TTGGTTGGTTGACTGGCTGGTTGGTTGGCTGGCTGGCTGGTTGGTTGGTTGGTTGGCTGG 260

Qy        589 ACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 648
                |   |  | | |  || | |   || || | || |  | |   | ||| | || |
Db        259 TTGGTAGGTTAGTTGGCTGGTTGGTTGGTTGGTTGGTTAGTTGGTTGACTGGTTGGTTAG 200

Qy        649 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGC 708
              | |   | ||| | || || | |   || || | || || | |   || |  | || ||
Db        199 TTGGTTGACTGGTTGGTTGGTTGACTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTGG 140

Qy        709 TAGTAGGGCTG 719
              | |   || ||
Db        139 TTGGTTGGTTG 129


RESULT 35
BZ236616
LOCUS       BZ236616                736 bp    DNA     linear   GSS 12-OCT-2002
DEFINITION  CH230-251M10.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-251M10, genomic survey sequence.
ACCESSION   BZ236616
VERSION     BZ236616.1  GI:23895644
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 736)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-251M10.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 251  row: M  column: 10
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .736
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-251M10"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      143 a    168 c    196 g    229 t
ORIGIN

  Query Match             11.4%;  Score 114.2;  DB 29;  Length 736;
  Best Local Similarity   72.0%;  Pred. No. 2.2e-11;
  Matches  149;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        371 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 430
              ||| || |||| || ||  |||| | | || |||||||||| || ||| || |||||||||
Db         16 GCTAGCCTGCTGGTGGGCTTGCTAGGTTGCTTGCTTGCTTCCTTGCTTGCTTGCTTGCTT 75

Qy        431 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              ||| ||| || |||||||||||| |||  | | ||||||| || |||  | ||||| |||
Db         76 GCTTGCTTGCTTGCTTGCTTGCTTGCTTTCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTT 135

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              |||  || || | |||||| || |||  | ||||| ||||||  || || | ||||||
Db        136 GCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTTGCTT 195

Qy        551 AGGGGACAGTAGGGCTGCTTGCTTGCT 577
                | | |  ||   |||||||  ||
Db        196 GCTTGCCTGATGGCTAGCTTGCTATCT 222


RESULT 36
BH604586
LOCUS       BH604586                633 bp    DNA     linear   GSS 15-DEC-2001
DEFINITION  BOHCF82TF BOHC Brassica oleracea genomic clone BOHCF82, genomic
            survey sequence.
ACCESSION   BH604586
VERSION     BH604586.1  GI:17857032
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 633)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOHCF82TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .633
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHCF82"
                     /clone_lib="BOHC"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
```

```
                genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       80 a    207 c    212 g    134 t
ORIGIN

  Query Match             11.4%;  Score 113.8;  DB 28;  Length 633;
  Best Local Similarity   51.7%;  Pred. No. 2.6e-11;
  Matches  315;  Conservative    0;  Mismatches  282;  Indels   12;  Gaps    2;

Qy        243 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAG 302
              || ||||  |     |    || |||| | || ||   | || ||||  |     |
Db          7 GCGTGCTCGGTCCCTACGCAGCGTGCTCGGTCCCTACGCAGCGTGCTCGGTCCCTACGCA 66

Qy        303 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 362
              || |||| | |       |||   ||  | |  ||  |  || |||| | |       ||
Db         67 GCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCACTACCCAGCGTGCTCGGTCGCTACGCA 126

Qy        363 GTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 422
              |  |   | | |||    ||   |   | |||||   | || |||| | |||||   |
Db        127 GCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCA 186

Qy        423 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 482
              || |||| | |||||   | || |||| | |||||   | || |||| | |||||   |
Db        187 GCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCA 246

Qy        483 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 542
              || || | | |||||   | || |||| | |||||   | || |||| | |||||   |
Db        247 GCGTGTTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCA 306

Qy        543 GCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATG 602
              || ||||  |  | |    | |||||| | |||| |  |   |  ||    ||
Db        307 GCGTGCTCGGTCGCAACGCAGCGTGCTTGGTCGCTACGTAGCGTGCTCGGTCCCC----- 361

Qy        603 GCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 662
                  ||    |  |||| ||   |||    |  |||| ||  ||||    |  ||||
Db        362 ------TACGTAGCGTGCTCGGTCCCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTC 415

Qy        663 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTG 722
              ||  ||||    |  |||| ||  ||||    |  | || ||  || |    |  ||||
Db        416 GGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCAACGTAGCGTGCTC 475

Qy        723 GGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 782
              ||  || |    |  | || || |||||    |  |||  |||     ||||  |
Db        476 GGTCGCAACGTAGCGTGCTCGGTTGCTATGTAGCGTGCACGGCGCTACGTAGCACGTAGC 535

Qy        783 GGCTGCTAG-TAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCT 841
              | | |   | | |  | ||| ||  ||||    |  |||  | | |||   | || ||||
Db        536 GACCGAGCGATCGTCCCGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCT 595

Qy        842 TGCTAGTGG 850
              | | | ||
Db        596 CGGTCGCGG 604


RESULT 37
BH328866
LOCUS       BH328866                740 bp    DNA     linear   GSS 03-DEC-2001
DEFINITION  CH230-93P12.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
            CH230-93P12, genomic survey sequence.
ACCESSION   BH328866
VERSION     BH328866.1  GI:17259580
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 740)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 EcoRI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-93P12.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 93  row: P  column: 12
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .740
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-93P12"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 1"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      181 a    199 c    149 g    211 t
ORIGIN

  Query Match             11.3%;  Score 113.4;  DB 28;  Length 740;
  Best Local Similarity   77.1%;  Pred. No. 3.1e-11;
  Matches  138;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy        371 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 430
              ||| || |||| || ||  |||| | | || ||||||||||||| ||| || |||||||||
Db        524 GCTAGCCTGCTGGTGGGCTTGCTAGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 583

Qy        431 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              ||  || || |||||||||||| ||| || ||||| |||||| ||| || |||||||||
Db        584 TCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTCCTTGCTTGCTTGCTTGCTT 643

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              || ||| || ||||| ||||||  || || |||||||  || | ||| || ||||  ||
Db        644 TCTTGCTTGCTTGCTTCCTTGCTTCCTTGCTTGCTTGCCTGATGGCTAGCTTGCTATCT 702


RESULT 38
CA047817
LOCUS       CA047817                712 bp    mRNA    linear   EST 04-MAR-2003
DEFINITION  ssalbrh012052 head Salmo salar cDNA, mRNA sequence.
ACCESSION   CA047817
VERSION     CA047817.1  GI:24354003
KEYWORDS    EST.
SOURCE      Salmo salar (Atlantic salmon)
  ORGANISM  Salmo salar
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
REFERENCE   1  (bases 1 to 712)
  AUTHORS   GRASP Consortium, Davidson,W.S., Koop,B.F. and
            http://web.uvic.ca/cbr/grasp.
  TITLE     A survey of Salmo salar transcripts from high complexity cDNA
            libraries
  JOURNAL   Unpublished
COMMENT     Contact: Koop BF
            Centre for Biomedical Research
            University of Victoria
            PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
            Tel: 250 472 4067
            Fax: 250 472 4075
```

```
            Email: bkoop@uvic.ca
            Centre for Biomedical Research, University of Victoria
            cDNA preparation and sequencing:
            Roberto Alberto, Marianne Beetz-Sargent, Maura Busby,
            Peter Hunt, Linda McKinnel, BF Koop.
            bioinformatics:
            Gordon D Brown
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .712
                     /organism="Salmo salar"
                     /mol_type="mRNA"
                     /strain="McConnell"
                     /db_xref="taxon:8030"
                     /clone_lib="head"
                     /note="Vector: pBlueScriptIISK+; Library Creator:
                     Kristian R von Schalburg ; Atlantic salmon tissue
                     contributors: Carlo Biagi, Mitch Uh and Robert Devlin (DFO
                     , Vancouver, B.C.), Simon Jones (PBS, Nanaimo, B.C.),
                     Seaspring Hatchery (Crofton, B.C.), Rachel Roper
                     (University of Victoria)"
BASE COUNT      168 a    159 c    194 g    191 t
ORIGIN

  Query Match            11.3%;  Score 113.2;  DB 14;  Length 712;
  Best Local Similarity   56.8%;  Pred. No. 3.3e-11;
  Matches  208;  Conservative    0;  Mismatches  158;  Indels    0;  Gaps    0;

Qy        222 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 281
              | |  ||  ||| |||| |  ||| |||||  ||   | ||||  ||| ||| ||||||
Db        337 TACTGGCTGGCTGGCTGACTGGCTGGCTGACTGGCTGGCAGGCTGGCTGGCTGGCTGGCT 396

Qy        282 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 341
                ||  |||   |    |  |||   ||  ||||  ||   | ||||  ||  ||| |||
Db        397 GACTAACTGGCTGACTGGCTGGCTGACTGACTGACTGGCTGGCAGGCTGGCAGGCTGGCT 456

Qy        342 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 401
              |||  |||  |||   |       ||    ||| ||  || |  |     || ||||||
Db        457 GGCTGGCTGACTGGCTGACTGACTGGCTGACTGACTGACTGGCTGACTGACTGGCTGGCT 516

Qy        402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461
               |||  ||  || ||||||  |||  ||  ||  ||| |   || |||  ||  |||||
Db        517 GGCTGACTGACTGGCTGGCTGGCTGACTAACTGACTGACTGACTGGCTGACTAACTGGCT 576

Qy        462 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 521
               ||| ||| ||| ||||||  |||  ||  ||  ||| |  ||| |||  ||  |||||
Db        577 GGCTGGCTGGCTGGCTGGCTGGCTGACTAACTGACTGACTGGCTGGCTGACTGACTGGCT 636

Qy        522 TGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGG 581
               |||  ||  || ||||||  |||  ||||  ||   |  |     ||  ||  || |
Db        637 GGCTGACTGACTGGCTGGCTGGCTGACTGACTGGCTGGCTGACTGACTAACTGACTGACT 696

Qy        582 GGACGG 587
              |   ||
Db        697 GACTGG 702


RESULT 39
AZB00556/c
LOCUS       AZB00556                 465 bp    DNA     linear   GSS 16-FEB-2001
DEFINITION  2M0058F10R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0058F10 R, genomic survey sequence.
ACCESSION   AZB00556
VERSION     AZB00556.1  GI:12952820
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 465)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0058  row: F  column: 10
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 465.
FEATURES             Location/Qualifiers
     source          1. .465
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0058F10"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      172 a     98 c     95 g     99 t      1 others
ORIGIN

  Query Match            11.1%;  Score 110.8;  DB 28;  Length 465;
  Best Local Similarity   71.8%;  Pred. No. 9e-11;
  Matches  145;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        332 CTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTG 391
              | ||   |||  | |  ||  | |  ||| |        || ||| || ||  |  |  ||
Db        202 CATGGGAGCTCTCTTAATTAATCATTGGATAAATTCATTGATGGATTCCTTATTTGTTTG 143

Qy        392 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 451
              || ||| || |||||||||||| ||| || ||| |||||||| ||| || ||| ||||||
Db        142 CTTGCTTGCTTGCTTGCTTGCTTGCTTGCATGCTTGCTTGCTTGCTTGCATGCTTG 83

Qy        452 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 511
              |  ||| || ||||||||||| ||| || ||||||||||  ||| || ||||||||||
Db         82 CATGCTTGCTTGCTTGCTTGCTTGCATGCTTGCTTGCATGCTTGCTTGCTTGCTTG 23

Qy        512 CTCGCTGGCCTGCTTGCTTGCT 533
              || ||| || ||| ||||||||
Db         22 CTTGCTTGCTTGCATGCTTGCT 1


RESULT 40
BH515818
LOCUS       BH515818                 564 bp    DNA     linear   GSS 13-DEC-2001
```

```
DEFINITION  BOHAC43TR BOHA Brassica oleracea genomic clone BOHAC43, genomic
            survey sequence.
ACCESSION   BH515818
VERSION     BH515818.1  GI:17723908
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 564)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOHAC43TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .564
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHAC43"
                     /clone_lib="BOHA"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       56 a    176 c    198 g    134 t
ORIGIN
```

```
  Query Match             11.0%;  Score 110.6;  DB 28;  Length 564;
  Best Local Similarity   51.3%;  Pred. No. 9.9e-11;
  Matches  287;  Conservative    0;  Mismatches  264;  Indels    8;  Gaps    1;

Qy        307 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAG 366
              || ||||  |  |  |    || |||| | |||||     || ||||  |  |  |
Db          1 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 60

Qy        367 GGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 426
              |  |||| | | ||||    |  ||||||  | ||    | || ||||||  | ||    |
Db         61 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 120

Qy        427 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 486
              || ||||||  | ||    | || ||||||  | ||    | || ||||||  | ||    |
Db        121 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 180

Qy        487 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 546
              || ||||||  | ||    | || ||||||  | ||    | || ||||||  | ||    |
Db        181 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTA 240

Qy        547 GCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTT 606
              ||      |   |    |  || |||| | |       ||  || | | |
Db        241 GCGTGCTCGGTCGCTACGTAGCGTGCTCGGT--------CGCTACGTAGCGTGCTCGGTC 292

Qy        607 GATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 666
              | ||    |   ||| ||  ||||    |   ||| ||  ||||    |   ||| ||
Db        293 GCTACGTAGCGCGCTCGGTCGCTACGTAGCGCGCTCGGTCGCTACGTAGCGCGCTCGGTC 352

Qy        667 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCT 726
              ||||    |   | | ||  ||||    |    || ||  ||||    |   ||| ||
Db        353 GCTACGTAGCGCGTTCGGTCGCTACGTAGCGCGCTCGGTCGCTACGTAGCGCGCTCGGTC 412

Qy        727 GCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 786
              ||||    |    || ||  ||||    |  |||| ||  ||||    |   |||| ||
Db        413 GCTACGTAGCGCGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTC 472

Qy        787 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTA 846
              ||||    |  |||| ||  ||||    |  |||  | | |||   | || |||| | |
Db        473 GCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTC 532

Qy        847 GTGGGGCCGCTTGCCTGCT 865
              |    |  || |||  | |
Db        533 GCTACGTAGCGTGCTCGGT 551
```

```
RESULT 41
BH443125
LOCUS       BH443125                490 bp    DNA     linear   GSS 12-DEC-2001
DEFINITION  BOHRJ25TR BOHR Brassica oleracea genomic clone BOHRJ25, genomic
            survey sequence.
ACCESSION   BH443125
VERSION     BH443125.1  GI:17628839
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOHRJ25TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .490
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHRJ25"
                     /clone_lib="BOHR"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       55 a    148 c    168 g    119 t
ORIGIN
```

```
  Query Match             10.9%;  Score 109;  DB 28;  Length 490;
  Best Local Similarity   53.4%;  Pred. No. 1.9e-10;
  Matches  256;  Conservative    0;  Mismatches  215;  Indels    8;  Gaps    1;

Qy        373 TGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 432
              | |||   |||   |   | |||||     || |||| | |||||     || |||| |
Db          2 TCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGG 61

Qy        433 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 492
              |||||     | | |||| | |||||     || |||| | |||||     || |||| |
Db         62 TCGCTACGTTACATGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGG 121

Qy        493 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAG 552
              |||||     || |||| | |||||     || |||| | |||||     || |||| |
Db        122 TCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGG 181

Qy        553 GGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACG--------CCTGCCTGATGGC 604
              |  |   |  |||| | | |||| |  |   |  ||          || || | |
Db        182 TCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGG 241

Qy        605 TTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 664
              | | ||    |  |||| ||  ||||    |  |||| ||  ||||    |  |||| ||
Db        242 TCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGG 301
```

```
Qy        665 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGG 724
              ||||    |  |||| ||  ||||    |  | || ||  ||||    |  |||| ||
Db        302 TCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGG 361

Qy        725 CTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 784
              ||||    |  | || ||  ||||    |  |||| ||  ||||    |  |||| ||
Db        362 TCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGACGCTACGTAGCGTGCTCGG 421

Qy        785 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTG 843
              ||||    |  |||| ||  ||||       |||  |   |||   | || |||| |
Db        422 ACGCTACGTAGCGTGCTCGGACGCTACGTACCGTGCTCGGACGCTACGTAGCGTGCTCG 480


RESULT 42
CNS04GRC
LOCUS       CNS04GRC                462 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            109C01 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL290001
VERSION     AL290001.1  GI:8028581
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 462)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .462
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="109C01"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG109AB01LP1~end : T7"
BASE COUNT       27 a      7 c    249 g    159 t     20 others
ORIGIN

  Query Match           10.9%;  Score 108.8;  DB 29;  Length 462;
  Best Local Similarity   52.5%;  Pred. No. 2.1e-10;
  Matches  208;  Conservative    9;  Mismatches  179;  Indels    0;  Gaps    0;

Qy        397 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 456
              | |  || |||  || | | | |  ||  || ||| | | : |  || :|| :|| : |

Db         58 TTGGATGGTTGGGTGGTTGGTTGGATGGNTGGTTGGTTGGBTGGKTGGKTGGBTGGKTGG 117

Qy        457 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 516
              | |  || |||  ||   | | |  ||  ||  || | | | |  || ||| ||| | |
Db        118 TTGGGTGGTTGGGTGGGTGGTTGGGTGGNTGGNTGGTTGGTTGGNTGGTTGGTTGGTTGG 177

Qy        517 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGC 576
              | |  || ||| ||| | | | |  || ||| || | ||   |  |||  | | | | |
Db        178 TTGNTTGGTTGGTTGGTTGGTTGGTTGGTTGNTGGGTGGGTGGGTGGGTGGGTGGGTGGG 237

Qy        577 TAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGG 636
              |  | ||  ||   ||  |   | ||| | | | |  |||  | |||| :| | |   ||
Db        238 TGGGTGGGTGGGTGGCTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTKGGTGGGTTGG 297

Qy        637 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 696
               || |||   | | || ||   | |||:  | | |: ||:  | |||   | | || ||
Db        298 TTGGTGGTTGGTTKGTTGGTTGGTTGGKTGGKTGGKTGGKTGGTTGGTTGGKTGGTTGGT 357

Qy        697 CTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGG 756
                  ||||  | | |  ||| | |||| || | |  ||     |||| || | |  ||
Db        358 TGGTTGGGTGGTTGGGTGGGTGGGTGGGTTGGTGGGTGGTTGGGTGGGTTGTTGGCTGGC 417

Qy        757 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGT 792
              |:|||   ||| || |||  |||||   ||| ||
Db        418 TGGYTGGCTGGCTGGTTGGGTGGCTGGTTGGCTGGT 453


RESULT 43
BH474846/c
LOCUS       BH474846                660 bp    DNA     linear   GSS 13-DEC-2001
DEFINITION  BOHLI64TF BOHL Brassica oleracea genomic clone BOHLI64, genomic
            survey sequence.
ACCESSION   BH474846
VERSION     BH474846.1  GI:17682957
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 660)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOHLI64TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .660
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHLI64"
                     /clone_lib="BOHL"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      112 a    217 c    250 g     81 t
ORIGIN

  Query Match           10.8%;  Score 108.4;  DB 28;  Length 660;
  Best Local Similarity   50.9%;  Pred. No. 2.5e-10;
  Matches  289;  Conservative    0;  Mismatches  271;  Indels    8;  Gaps    1;

Qy        305 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 364
              | || |||| |  | ||    || |||| | ||||    | || ||||  |  | ||
```

```
Db        654 CAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACG 595

Qy        365 AGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 424

Db        594 CAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACG 535

Qy        425 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 484

Db        534 CAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACG 475

Qy        485 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 544

Db        474 CAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACG 415

Qy        545 TTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGC 604

Db        414 CAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGT--------CGCCACGCAGCGTGCTCGG 363

Qy        605 TTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 664

Db        362 TCGCCACGCAGCGTGCCCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCCCGG 303

Qy        665 CTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGG 724

Db        302 TCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGG 243

Qy        725 CTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 784

Db        242 TCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGG 183

Qy        785 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGC 844

Db        182 TCGCCACGCAGCGTGCTCGGTCGCCATGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGG 123

Qy        845 TAGTGGGGCCGCTTGCCTGCTACTAGGG 872

Db        122 TCGCCACGCAGCGTGCTCGGTCGTAGCG 95


RESULT 44
BZ060388/c
LOCUS       BZ060388                732 bp    DNA     linear   GSS 10-OCT-2002
DEFINITION  l1g98b12.g1 B.oleracea002 Brassica oleracea genomic, genomic survey
            sequence.
ACCESSION   BZ060388
VERSION     BZ060388.1  GI:23669293
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 732)
  AUTHORS   Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash
            ,W., Rabinowicz,P.D. and Wilson,R.K.
  TITLE     Whole genome shotgun reads from Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: l1g98  row: b  column: 12
            Seq primer: -21UPpOT forward
            Class: shotgun
            High quality sequence start: 8
            High quality sequence stop: 463.
FEATURES             Location/Qualifiers
     source          1. .732
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3712"
                     /clone_lib="B.oleracea002"
                     /note="Vector: pOTw13; Whole genome shotgun library from
                     flowering buds. DNA was purified  from a crude nuclear
                     prep using Brassica oleracea TO1000DH3 buds  provided by
                     Thosmas Osborn at the University of Wisconsin. Genomic
                     DNA was provided by Pablo Rabinowicz (CSHL) and the
                     shotgun library  prepared at Washington University Genome
                     Sequencing Center."
BASE COUNT      134 a    226 c    274 g     97 t      1 others
ORIGIN

  Query Match            10.8%;  Score 108.4;  DB 28;  Length 732;
  Best Local Similarity  50.9%;  Pred. No. 2.5e-10;
  Matches  289;  Conservative    0;  Mismatches  271;  Indels    8;  Gaps    1;

Qy        305 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 364

Db        723 CAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACG 664

Qy        365 AGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 424

Db        663 CAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACG 604

Qy        425 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 484

Db        603 CAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACG 544

Qy        485 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 544

Db        543 CAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACG 484

Qy        545 TTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGC 604

Db        483 CAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGT--------CGCCACGCAGCGTGCTCGG 432

Qy        605 TTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 664

Db        431 TCGCCACGCAGCGTGCCCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCCCGG 372

Qy        665 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGG 724

Db        371 TCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGG 312

Qy        725 CTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 784

Db        311 TCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGG 252

Qy        785 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGC 844

Db        251 TCGCCACGCAGCGTGCTCGGTCGCCATGCAGCGTGCTCGGTCGCCACGCAGCGTGCTCGG 192

Qy        845 TAGTGGGGCCGCTTGCCTGCTACTAGGG 872

Db        191 TCGCCACGCAGCGTGCTCGGTCGTAGCG 164


RESULT 45
BH725237
LOCUS       BH725237                821 bp    DNA     linear   GSS 20-FEB-2002
DEFINITION  BOMHG42TF BO_2_3_KB Brassica oleracea genomic clone BOMHG42,
            genomic survey sequence.
ACCESSION   BH725237
VERSION     BH725237.1  GI:18830632
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 821)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOMHG42TR
            Contact: Chris Town
```

```
      TIGR
      9712 Medical Center Drive, Rockville, MD 20850, USA.
      Tel: 301-838-3523
      Fax: 301-838-0208
      Email: cdtown@tigr.org
      DNA is from a doubled haploid provided by Tom Osborn.
      Seq primer: TF
      Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .821
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOMHG42"
                     /clone_lib="BO_2_3_KB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       86 a    284 c    280 g    171 t
ORIGIN

  Query Match           10.7%;  Score 107.4;  DB 28;  Length 821;
  Best Local Similarity   49.1%;  Pred. No. 3.8e-10;
  Matches  318;  Conservative    0;  Mismatches  321;  Indels    8;  Gaps    1;

Qy        227 GCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTT 286
              ||  | |||||     |||||||  |  |  |    || |||| | |||||     || |
Db         14 GCTCGGTCGCTTCATAGCTTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGT 73

Qy        287 GCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 346
              |||  |  |  |    || |||| | |        ||   ||| | | |||   | || |
Db         74 GCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGT 133

Qy        347 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 406
              ||  | |         ||   | | | | |||   |||   | | | |||||     || |
Db        134 GCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGT 193

Qy        407 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 466
              ||  | |||||     || |||  | |||||     || |||  | || ||     || |
Db        194 GCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCTCTACGTAGCGT 253

Qy        467 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 526
              ||  | |||||     || |||  | |||||     || |||  | |||||     || |
Db        254 GCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGT 313

Qy        527 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACG 586
              ||  | |||||     || |||   |  |  |     |  |||  | | |||| |  |
Db        314 GCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGT 373

Qy        587 GTACGCCTGCCTGATGGCTTGA--------TAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 638
              |  ||   ||    | || ||         ||    |  |||  ||  ||||    |  |
Db        374 GCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGT 433

Qy        639 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 698
              ||  ||  ||||    |  |||  ||   |||    |  |||  ||  ||||    |  |
Db        434 GCCCGGTCGCTACGTAGCGTGCCCGGTCACTACGTAGCGTGCCCGGTCGCTACGTAGCGT 493

Qy        699 CCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCT 758
               |  ||  ||||    |  |||  ||  ||||    |  |  |  ||  ||||    |  |
Db        494 GCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGT 553

Qy        759 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 818
              ||  ||  ||||    |  |||  ||  ||||    |  |||  ||  ||||    |  |
Db        554 GCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGT 613

Qy        819 GCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCT 865
              |||  |  | |||   |  ||  |||  |  |  |    |  ||  |||| | |
Db        614 GCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGT 660


RESULT 46
AI676341
LOCUS       AI676341                438 bp    mRNA    linear   EST 19-MAY-1999
DEFINITION  etmEST0088 EtH1 Eimeria tenella cDNA clone etmb035 5', mRNA
            sequence.
ACCESSION   AI676341
VERSION     AI676341.1  GI:4876821
KEYWORDS    EST.
SOURCE      Eimeria tenella
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 438)
  AUTHORS   Wan,K.L., Chong,S.P., Ng,S.T., Tomley,F.M. and Jangi,M.S.
  TITLE     Survey of genes in Eimeria tenella merozoite
  JOURNAL   Unpublished
COMMENT     Contact: Wan KL
            Centre for Gene Analysis and Technology
            Universiti Kebangsaan Malaysia
            43600 UKM Bangi, Selangor DE, Malaysia
            Tel: 6 03 8292997
            Fax: 6 03 8293249
            Email: klwan@pkrisc.cc.ukm.my
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7
            Seq primer: SK.
FEATURES             Location/Qualifiers
     source          1. .438
                     /organism="Eimeria tenella"
                     /mol_type="mRNA"
                     /strain="Houghton"
                     /db_xref="taxon:5802"
                     /clone="etmb035"
                     /dev_stage="second generation merozoite"
                     /lab_host="XL1-Blue MRF'"
                     /clone_lib="EtH1"
                     /note="Vector: Lambda ZAPII; Site_1: EcoRI; Site_2: XhoI;
                     Second generation merozoites of E. tenella H were purified
                     by column  chromatography and mRNA extracted using a
                     FASTTRACK kit (Invitrogen).  cDNA was synthesised and a
                     Uni-ZAP XR library was constructed using  cDNA synthesis
                     kit, ZAP-cDNA synthesis kit and ZAP-cDNA Gigapack III
                     Gold cloning kit (Stratagene). The library was amplified
                     once through  E. coli XL1-Blue MRF'."
BASE COUNT       87 a    121 c    150 g     80 t
ORIGIN

  Query Match           10.7%;  Score 106.8;  DB 9;  Length 438;
  Best Local Similarity   61.5%;  Pred. No. 4.8e-10;
  Matches  171;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy        366 GGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 425
              || |  | ||   ||| | |  |      |||||   || |  |  |  |||||||  |||
Db        121 GGCCCACAGGGCAGCTGGGACTGGCTGGCGCTGAGGTGGTCTCGGGACCGCTGGCTAGCT 180

Qy        426 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 485
              |||  ||| ||| ||  |||||||  | | |||||   ||| ||| ||| ||||||  |||
Db        181 AGCTAGCTAGCTAGCTAGCTTGCTGGATAGCTGGATAGCTGGCTAGCTGGCTGGCTAGCT 240

Qy        486 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 545
              | | |||  ||| ||  |||  ||| ||||||||||  |||  |||  ||| ||| ||  |||
Db        241 GGGTCGCTGGCTAGCTAGCTGGCTAGCTCGCTGGCTAGCTAGCTGGCTAGCTAGCTGGCT 300

Qy        546 TGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCT 605
              |||    ||  ||  ||   ||| ||| ||||    | |   | ||  ||  | |||||
Db        301 AGCTAGCTGGCTAGCTGGCTGGCTAGCTAGCTAGCTAGCCACCAAGCTAGCTAGCTGGCT 360

Qy        606 TGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 643
              |  ||| |   ||| |  | |||||  ||   ||| |
Db        361 AGCGAGTTGCCTTGCCGCCCCGCTAGGTGGCTAGCTAG 398


RESULT 47
BZ263814
LOCUS       BZ263814                711 bp    DNA     linear   GSS 15-OCT-2002
```

```
DEFINITION  CH230-420O12.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-420O12, genomic survey sequence.
ACCESSION   BZ263814
VERSION     BZ263814.1  GI:23975613
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 711)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-420O12.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 420  row: O  column: 12
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .711
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-420O12"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      149 a    157 c    191 g    214 t
ORIGIN

  Query Match             10.6%;  Score 106;  DB 29;  Length 711;
  Best Local Similarity   62.4%;  Pred. No. 6.9e-10;
  Matches  166;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy        341 TGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGC 400
              | ||  || ||||| |||   |   || |||| ||||||| |   |  |||| ||| ||
Db         21 TAGCTAGCCTGCTGGTGGGCTTGCTAGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 80

Qy        401 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 460
              |||||||||||| ||| || ||||| ||||||  || || | ||||||| || |||  |
Db         81 TTGCTTGCTTGCTTGCTTGCTTGCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTTGCTTCC 140

Qy        461 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 520
              |||||  ||||||  || || ||||||| || | ||| || ||||  ||| ||  ||  |
Db        141 TTGCTTCCTTGCTTCCTTGCTTGCTTGCCTGATGGCTAGCTTGCTATCTTCCTTCCTTCC 200

Qy        521 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAG 580
              | ||| ||| ||  ||  | | ||| ||       |        |||| ||| ||
Db        201 TTCCTTTCTTCCTTCCTTCCTTCCTTCCTTTCCTTCTTTATTGCTTTCTTTCTTTC 260

Qy        581 GGGACGGTACGCCTGCCTGATGGCTT 606
                  |    | | | | | ||  |||
Db        261 TTTCCTTCTCCCTTCCTTCATTCCTT 286
```

```
RESULT 48
BH700477
LOCUS       BH700477                528 bp    DNA     linear   GSS 20-FEB-2002
DEFINITION  BOMLB36TF BO_2_3_KB Brassica oleracea genomic clone BOMLB36,
            genomic survey sequence.
ACCESSION   BH700477
VERSION     BH700477.1  GI:18776991
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOMLB36TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOMLB36"
                     /clone_lib="BO_2_3_KB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       81 a    138 c    178 g    131 t
ORIGIN

  Query Match             10.5%;  Score 105.4;  DB 28;  Length 528;
  Best Local Similarity   56.4%;  Pred. No. 8.7e-10;
  Matches  224;  Conservative    0;  Mismatches  161;  Indels   12;  Gaps    1;

Qy        388 GCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 447
              |||| || ||     || |||| | |||||     || ||||  |||||     || ||
Db        143 GCTGGTCACTACGTAGCGTGCTCGGTCGCTACGTGGCGTGCTCTGTCGCTACGTGGCGTG 202

Qy        448 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 507
              || | |||||     || |||| | |||||     || |||| | |||||     |||||
Db        203 CTCGGTCGCTACGTGGCGTGCTCGGTCGCTACATGGCGTGCTCGGTCGCTACGTGGCTTG 262

Qy        508 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTG 567
              || | |||||     || |||| | |||||     || ||||  |  |     ||  ||
Db        263 CTCGGTCGCTACGTGGCGTGCTCGGTCGCTACGTGGCGTGCTCGGTCGCTCCGTGGCGTG 322

Qy        568 CTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTG 627
              || | | |||| | |            || || | | | | | ||   ||  |||| ||  |
Db        323 CTCGGTCGCTACGTG-----------GCGTGCTCGGTCGCTACGTGGCGTGCTCGGTCG 370

Qy        628 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTG 687
              |||   ||  |||| ||  ||||   ||  || | ||  ||||    |  || | ||  |
Db        371 CTACGTGGCGTGCTCGGTCGCTACGTGGCGTGTTCGGTCGCTACGTTGCGTGTTCGGTCG 430

Qy        688 CTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTG 747
              |||   ||  | || ||  ||||   ||  |||| ||  ||||   ||  | || ||  |
Db        431 CTACGTGGCGTGCTCGGTCGCTACGTGGCGTGCTCGGTCGCTACGTGGCGTGCTCGGTCG 490

Qy        748 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 784
              |||   ||  |||| ||  ||||   ||  |||||||
Db        491 CTACGTGGCGTGCTCGGTCGCTACGTGGCGTGCTGGG 527
```

```
RESULT 49
CB504454/c
LOCUS       CB504454                710 bp    mRNA    linear   EST 16-MAY-2003
DEFINITION  ssalplnb502144_rev gut Salmo salar cDNA, mRNA sequence.
ACCESSION   CB504454
VERSION     CB504454.1  GI:29315680
KEYWORDS    EST.
SOURCE      Salmo salar (Atlantic salmon)
  ORGANISM  Salmo salar
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
REFERENCE   1  (bases 1 to 710)
  AUTHORS   GRASP Consortium, Davidson,W.S., Koop,B.F. and
            http://web.uvic.ca/cbr/grasp.
  TITLE     A survey of Salmo salar transcripts from high complexity cDNA
            libraries
  JOURNAL   Unpublished
COMMENT     Contact: Koop BF
            Centre for Biomedical Research
            University of Victoria
            PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
            Tel: 250 472 4067
            Fax: 250 472 4075
            Email: bkoop@uvic.ca
            Genome Sciences Centre, BC Cancer Agency cDNA preparation,
            sequencing and bioinformatics: Y Butterfield, R Kirkpatrick, J
            Asano, N Girn, R Guin, D Lee, S Lee, T Olson, P Pandoh, A Prahbu, D
            Smailus, L Spence, J Stott, S Taylor, G Yang, J Schein, S Jones and
            M Marra.
FEATURES             Location/Qualifiers
     source          1. .710
                     /organism="Salmo salar"
                     /mol_type="mRNA"
                     /strain="McConnell"
                     /db_xref="taxon:8030"
                     /clone_lib="gut"
                     /note="Vector: pBlueScriptIISK+; Library Creator:  Matthew
                     L Rise ; Atlantic salmon tissue contributors: Carlo Biagi,
                     Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.), Simon
                     Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery (Crofton,
                     B.C.), Rachel Roper (University of Victoria)"
BASE COUNT      172 a    210 c    184 g    144 t
ORIGIN

  Query Match           10.5%;  Score 105.4;  DB 14;  Length 710;
  Best Local Similarity   60.1%;  Pred. No. 8.8e-10;
  Matches  175;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        261 AGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGA 320
              ||   | || ||| |||| |  ||| ||||   ||   |  |||  ||| ||||   |
Db        692 AGCTTTACTGGCTGGCTGACTGGCTGGCTGGCTGGCAGGCTGGCTGGCTGGCTGGCTGAC 633

Qy        321 CAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGC 380
              |   |||  ||| ||| |||||| ||| ||||   ||   |  |     ||| ||  |
Db        632 TAACTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGACTAACTGACTGAC 573

Qy        381 TAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 440
              |    ||   ||  |||||  ||| |||  || |||| |  ||| |||  || |||| |
Db        572 TGACTGGCTGACTAACTGGCTGGCTGGCTGACTGGCTGACTGGCTGGCTGACTGGCTGAC 513

Qy        441 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 500
                ||  ||  || ||||||   || |||  ||  ||| |   || ||| ||| ||||||
Db        512 TAACTGACTGACTGGCTGGCTGACTGGCTGACTAACTGACTGACTGGCTGGCTGGCTGGC 453

Qy        501 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGA 551
                ||  || ||| ||||||  ||| ||| |||  |||||  ||| |||||
Db        452 TGACTAACTGGCTGGCTGGCTGGCTGGCTGACTGGCTGGCTGGCTGA 402


RESULT 50
BH483727
```

```
LOCUS       BH483727                569 bp    DNA     linear   GSS 13-DEC-2001
DEFINITION  BOGCW50TF BOGC Brassica oleracea genomic clone BOGCW50, genomic
            survey sequence.
ACCESSION   BH483727
VERSION     BH483727.1  GI:17691831
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 569)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOGCW50TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .569
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGCW50"
                     /clone_lib="BOGC"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       69 a    178 c    188 g    134 t
ORIGIN

  Query Match           10.5%;  Score 105;  DB 28;  Length 569;
  Best Local Similarity   52.6%;  Pred. No. 1e-09;
  Matches  254;  Conservative    0;  Mismatches  225;  Indels    4;  Gaps    1;

Qy        387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 446
              ||| |||   |  | |||| | | |||   | || |||| | | |||   | || ||||
Db          4 GGCCGCTACGTACCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTC 63

Qy        447 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 506
              | | |||   | || ||||  |   ||    | || ||||  | |  ||   | || ||||
Db         64 GGTCGCTACGTAGCGTGCTCGGCCGCAATGTAGCGTGCTCGGTCCCTACGTAGCGTGCTC 123

Qy        507 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 566
              . | |  ||   | || |||| | |  ||   | || |||| | |         ||   |
Db        124 GGTCCCTACGTAGCGTGCTCGGTCCCTACGTAGCGTGCTCGGTCCCTACGTAGCGTGCTC 183

Qy        567 GCTTGCTTGCTAAGGGGACGGTACG----CCTGCCTGATGGCTTGATAGTAGGGCTGCTG 622
              | | |||   ||  | |   | |||        || || | |   ||||    | ||||
Db        184 GGTCGCTACGTAGCGTGCTCGGACGATACGTAGCGTGCTCGGACGATACGTAGCGTGCTC 243

Qy        623 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 682
              ||  |||| . |  |||| ||  ||||    |  |||| ||  ||||    | ||||
Db        244 GGTCGCTACGTAGCATGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTC 303

Qy        683 GGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTG 742
              ||  ||||    |  | || ||  ||||    |  |||| ||  ||||    |  | ||
Db        304 GGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTC 363

Qy        743 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 802
              ||  ||||    |  |||| ||  ||||    |  |||| ||  ||||    |  ||||
Db        364 GGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTC 423

Qy        803 GGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCT 862
              ||  ||||    |  ||| | | |||   | || |||| | | |    |  || |||
Db        424 GGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTC 483
```

```
Qy        863 GCT 865
              | |
Db        484 GGT 486


RESULT 51
BH267859
LOCUS       BH267859                749 bp    DNA     linear   GSS 30-NOV-2001
DEFINITION  CH230-66L14.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
            CH230-66L14, genomic survey sequence.
ACCESSION   BH267859
VERSION     BH267859.1  GI:17180169
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 749)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 EcoRI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-66L14.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 66  row: L  column: 14
            Seq primer: SP6
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .749
                      /organism="Rattus norvegicus"
                      /mol_type="genomic DNA"
                      /strain="BN/SsNHsd/MCW"
                      /db_xref="taxon:10116"
                      /clone="CH230-66L14"
                      /sex="Female"
                      /cell_type="Brain"
                      /clone_lib="CHORI-230 Segment 1"
                      /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                      CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                      Pieter de Jong"
BASE COUNT      175 a    208 c    135 g    231 t
ORIGIN

  Query Match             10.4%;  Score 104.6;  DB 28;  Length 749;
  Best Local Similarity   69.1%;  Pred. No. 1.2e-09;
  Matches  143;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy        371 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 430
              ||| || |||| || ||  |||| | | || ||||||||||||| ||| || |||||||||
Db        482 GCTAGCCTGCTGGTGGGCTTGCTAGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 541

Qy        431 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              |||  || || ||||| ||||||  || || | ||||||| || |||  | ||||| |||
Db        542 GCTTTCTTGCTTGCTTACTTGCTTCCTTGCTTCCTTGCTTCCTTGCTTACTTGCTTCCTT 601

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              |||  || || | ||||||| || ||| || ||| || | ||| ||| ||   ||| ||
Db        602 GCTTCCTTGCTTCCTTGCTTCCTTGCTTGCTTGCCTGATGGCTAGCTTGCTATCTTACTT 661

Qy        551 AGGGGACAGTAGGGCTGCTTGCTTGCT 577
              |   |          | ||| ||| ||
Db        662 ACTTGCTTACTTTCTTCCTTCCTTACT 688


RESULT 52
BH544637
LOCUS       BH544637                794 bp    DNA     linear   GSS 14-DEC-2001
DEFINITION  BOGGJ62TF BOGG Brassica oleracea genomic clone BOGGJ62, genomic
            survey sequence.
ACCESSION   BH544637
VERSION     BH544637.1  GI:17796418
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 794)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOGGJ62TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES              Location/Qualifiers
     source           1. .794
                      /organism="Brassica oleracea"
                      /mol_type="genomic DNA"
                      /strain="TO1000DH3"
                      /db_xref="taxon:3712"
                      /clone="BOGGJ62"
                      /clone_lib="BOGG"
                      /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                      genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      157 a    235 c    235 g    167 t
ORIGIN

  Query Match             10.4%;  Score 104.4;  DB 28;  Length 794;
  Best Local Similarity   53.6%;  Pred. No. 1.3e-09;
  Matches  241;  Conservative    0;  Mismatches  206;  Indels    3;  Gaps    1;

Qy        220 GCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGG 279
              || || |||  | |||||   | || ||||  |  |  |    || |||||| ||||
Db        348 GCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTTGGTCGCTAC 407

Qy        280 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCG 339
              | || ||||  |  |  |    || |||| | |       |||   ||  | | |||
Db        408 GCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTAC 467

Qy        340 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGG 399
              || |||| | |       |||   | | | | |||    ||   | | | |||||
Db        468 GCAGCGTGCTCGGTCGCTACGCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTAC 527

Qy        400 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 459
              | || ||| | |||||   | || |||  | |||||   | || |||  | |||||
Db        528 GCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTAC 587

Qy        460 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 519
              | || ||| | |||||   | || |||  | |||||   | || |||  | |||||
Db        588 GCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTAT 647

Qy        520 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAA 579
              | || |||  | |||||   | || |||   |  |  |    |  ||| | | ||||
Db        648 GCAGCGTGCCCGGTCGCTACGCAGCGTGCCCGGTCGCTACGCAGCATGCCCGGTCGCTAC 707
```

```
Qy        580 GGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 639
              |  |   |     | || || | | | | ||    |  ||||  ||  ||||    |  ||
Db        708 GCAGCATG---CTCAGCGTGCTCGGTCGCTACGCAGCGTGCTCGGTCGCTACGCAGCGTG 764

Qy        640 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 669
              ||  ||  ||||    |  ||||  ||  |||
Db        765 CTCGGTCGCTACGCAGCGTGCTCGGTCGCT 794


RESULT 53
BX405071/c
LOCUS       BX405071               1201 bp    mRNA    linear   EST 13-MAY-2003
DEFINITION  BX405071 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
            clone CS0DG006YG06 3-PRIME, mRNA sequence.
ACCESSION   BX405071
VERSION     BX405071.1  GI:30648111
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 10245.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0AG006BD03NP1&cluster=10245.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0AG006BD03NP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DG006YG06"
                     /tissue_type="B CELLS (RAMOS CELL LINE)"
                     /cell_line="RAMOS CELL LINE"
                     /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
BASE COUNT       62 a    448 c    462 g      5 t    224 others
ORIGIN

  Query Match           10.4%;  Score 104.2;  DB 13;  Length 1201;
  Best Local Similarity   40.4%;  Pred. No. 1.5e-09;
  Matches  218;  Conservative   58;  Mismatches  264;  Indels    0;  Gaps    0;

Qy          9 GGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCTCCGC 68
              |:|   |:|| :|| | ||  :| |::|: |:|| :  :  |      |  |  |    :
Db        917 GSCCSCGYGGCSCTGCMCCCTSCGCSSCKTCSCCGSCGSCGCGGCCGCGTGGCGCCAGCS 858

Qy         69 GCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGAC 128
              ||  |: |: |:|| |:||| || || |  ||  ||:||:|   | || :  ||||:  |
Db        857 GCGCCSCGSCGSGGCGSCGGCGCGCGCGCGGCGGCGSCGSGCGGCCGCCSCCGGCGSCGC 798

Qy        129 CGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGACG 188
              : ::  :||| |    ||:  :  ||::    ||:| | |:||  | :|    |:  :
Db        797 GSCSSCGSCCGCGCCCGCGSCCSSSCGSSSGCGCGSGSGCGSAGGCGTSGCKGTGSCGSC 738

Qy        189 GCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGC 248
              | |   |  | : | | |    :|||   |  ||:| ||   : |:|  |::  |  ||

Db        737 GGGGTSGGTGGGKSGSGGGCGCGSGGGCGTGCGTGSCGGCGCTSGCKCGCGSBCTCGCGC 678

Qy        249 TGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGC 308
              : |: |   |  :||  |:  || :||  ||  ||  |: | :| | : |   ||  |
Db        677 GSSGSCGCGCGCGSGCGCGSGCGCGSGCGCGCGCGCGSGGGSGTGGSGGGSCGCGCGT 618

Qy        309 TTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGG 368
              || : :|:   |  ::|  |   ||:|||  |:  ||  ||    | |   |   |
Db        617 GSGCGSTSGSTSGCGSKSKCGCGGGCGCKCGCGCGSGCGCGCGCGCGCGCGCGCGCGC 558

Qy        369 CTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 428
              :|  || :||  |   |   || |||  ||  ||  ||  || |||  ||  ||  ||
Db        557 GCKCGCGCGYGCGCGCGCGCGCGCGCGCGCKCGCGCGCGCGCGCGCGCGCGCGCGC 498

Qy        429 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 488
              || |||  || |||  ||  || |||  ||  ||  ||  |: :|   ||  ||  ||
Db        497 GCGCGCGCGCGCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGYGYGTGCGCGCGCGC 438

Qy        489 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 548
              || |||   |  ||  |  || | ||  |   |   |   | | ||  |   |  ||
Db        437 GCGCGCGCGCNCTCTCTCNCNCGCNCNCTCNCGCNCNCNCNCNCNCTCNCNCNCNCGC 378


RESULT 54
CB952830
LOCUS       CB952830               819 bp    mRNA    linear   EST 29-APR-2003
DEFINITION  AGENCOURT_13688077 NIH_MGC_176 Mus musculus cDNA clone
            IMAGE:30304742 5', mRNA sequence.
ACCESSION   CB952830
VERSION     CB952830.1  GI:30208948
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 819)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein
             cDNA Library Preparation: Michael Brownstein Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM77  row: f  column: 15
            High quality sequence stop: 212.
FEATURES             Location/Qualifiers
     source          1. .819
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30304742"
                     /lab_host="DH10B (T1-phage-resistant)"
                     /clone_lib="NIH_MGC_176"
                     /note="Organ: kidney; Vector: pDNR-LIB; Site_1: SfiI
                     (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc); cDNA made
                     by oligo-dT priming and directionally cloned. 5' and 3'
                     adaptors were used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Full-length
                     enriched library was constructed using the Clontech
                     Creator SMART kit and size-selected to contain the 0.5 kb
                     size fraction. Library created in the laboratory of M.
                     Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."
BASE COUNT       93 a    206 c    242 g    277 t      1 others
ORIGIN

  Query Match           10.4%;  Score 104;  DB 14;  Length 819;
  Best Local Similarity   56.0%;  Pred. No. 1.6e-09;
```

```
  Matches  197;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy        336 CTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCG 395
              ||   || | |   |    ||   ||   || ||   | ||||| | | || ||   ||| |
Db         29 CTTATTGCCTTTTTGATTGGTTGGTTGGTTGGTTGGTTGGCTGGTTTGTTGGCTGGCTGG 88

Qy        396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455
              |||   ||| | | | | |||||   ||| ||| ||| | ||||  | | | | | | |
Db         89 TTGGTTGGCTGGTTGGTTGGCTGGTTGGCTGGCTGGCTGGTTGGCTGGTTGGTTGGTTGG 148

Qy        456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515
              |||   | | | | | | | |||   ||| ||| ||| |||||   ||| | | | | |
Db        149 TTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGGCTGGCTGGTTGGCTGGTTGGTTGGTTGG 208

Qy        516 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTG 575
              |||   ||| | | | | ||||||  ||| | ||   ||   || ||   | | | | | |
Db        209 TTGGTTGGCTGGTTGGTTGGCTGGCTGGCTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGG 268

Qy        576 CTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGG 635
              |    ||  ||   |  ||  | ||| | ||| || ||  |||||   |    | ||
Db        269 TTGGTTGGCTGGCTGGATGGCTGGCTGGTTGGATGGTTGGTTGGCTGGCTGGTCGATTGG 328

Qy        636 GCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTG 687
              | |||| || ||| ||  ||||| | |||||   ||  || ||   || ||
Db        329 CCGGCTGCGCGGCTGGTTTGGCTGGTTGGCTGGCTGTTTGGTTGGCTGGTTG 380


RESULT 55
BH677755
LOCUS       BH677755                677 bp    DNA     linear   GSS 19-FEB-2002
DEFINITION  BOHYJ15TR BO_2_3_KB Brassica oleracea genomic clone BOHYJ15,
            genomic survey sequence.
ACCESSION   BH677755
VERSION     BH677755.1  GI:18748198
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 677)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .677
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHYJ15"
                     /clone_lib="BO_2_3_KB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      101 a    202 c    222 g    152 t
ORIGIN

  Query Match            10.2%;  Score 102.6;  DB 28;  Length 677;
  Best Local Similarity  51.9%;  Pred. No. 2.8e-09;
  Matches  280;  Conservative    0;  Mismatches  254;  Indels    5;  Gaps    2;

Qy        331 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCT 390
              |||   | ||  ||  | | |||  |  |   |   |||  ||   | ||  |   |||
Db        117 GCTACGTAGCATGCTCGGTCGCTACGTAGCGTGCTCGGCGACTACGTAGCGTGCTCGGCC 176

Qy        391 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 450
              |||   |  | |||| | | |||   | || |||| | | |||   | || |||| | |
Db        177 GCTACGTACCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTC 236

Qy        451 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 510
              |||   | || |||| |   ||    | || |||| | |   ||   | || |||| | |
Db        237 GCTACGTAGCGTGCTCGGCCGCAATGTAGCGTGCTCGGTCCCTACGTAGCGTGCTCGGTC 296

Qy        511 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTT 570
              ||   | || |||| | |  ||   | || |||| | |         ||   |  | |
Db        297 CCTACGTAGCGTGCTCGGTCCCTACGTAGCGTGCTCGGTCCCTACGTAGCGTGCTCGGTC 356

Qy        571 GCTTGCTAAGGGGACGGTACG----CCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCT 626
              |||   ||  | |   | |||       || || | |   ||||    |  |||| ||
Db        357 GCTACGTAGCGTGCTCGGACGATACGTAGCGTGCTCGGACGATACGTAGCGTGCTCGGTC 416

Qy        627 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 686
              ||||    |  |||| ||  ||||    |  |||| ||  ||||    |  |||| ||
Db        417 GCTACGTAGCATGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTC 476

Qy        687 GCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCT 746
              ||||    |  | || ||  ||||    |  |||| ||  ||||    |  | || ||
Db        477 GCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTC 536

Qy        747 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 806
              ||||    |  |||| ||  ||||    ||||||  ||  ||||    |  |||| ||
Db        537 GCTACGTAGCGTGCTCGGTCGCTA-CGTAGCTGCTCGGTCGCTACGTAGCGTGCTCGGTC 595

Qy        807 GCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCT 865
              ||||    |  |||  | | |||   | || |||| | | |    |  || |||  | |
Db        596 GCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGT 654


RESULT 56
CNS02WKU
LOCUS       CNS02WKU                881 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            176H16 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL217191
VERSION     AL217191.1  GI:7876010
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 881)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
```

```
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .881
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="176H16"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG176DD08SP1~end :
                     PUC-Ori"
BASE COUNT      104 a    217 c    166 g    311 t     83 others
ORIGIN

  Query Match             10.2%;  Score 102.6;  DB 29;  Length 881;
  Best Local Similarity   54.9%;  Pred. No. 2.9e-09;
  Matches  196;  Conservative    3;  Mismatches  158;  Indels    0;  Gaps    0;

Qy        304 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 363
              ||||||    |   | .||        |  :|:|  ||: |||||| ||       | | |
Db        522 CCTGCTGATTATTGTTACTACCTATTTTTHTRCCTGCYTGCCTGCCTGTCTGTCTGCCTG 581

Qy        364 TAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 423
                | ||||  || |||  ||  | ||||  |   |||||  ||  |||  ||  |||||
Db        582 ACTGCCTGCCTGCCTGCCTGTCTGTCTGCCTGACTGCCTGTCTGTCTGCCTGCCTGCCTG 641

Qy        424 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 483
              | ||| ||   |   | |||| ||  |||  |   | |||  ||| ||| ||  |||||
Db        642 CCTGCCTGTCTGTCTGTCTGCCTGACTGCCTGTCTGTCTGTCTGCCTGCCTGCCTGCCTG 701

Qy        484 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 543
              | ||| |||  |   | |||  ||| ||   ||  |||||  ||  |||  ||  |||||
Db        702 CCTGCCTGCCTGTCTGTCTGTCTGCCTGACTGCCTGCCTGTCTGTCTGCCTGCCTGCCTG 761

Qy        544 CTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGG 603
              | |||      | | ||  | |||  ||| ||       | | |   |||||||||   |
Db        762 CCTGCCTGTCTGTCTGTCTGCCTGACTGCCTGTCTGTCTGTCTGCCTGCCTGCCTGCCTG 821

Qy        604 CTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 660
              | ||   ||  | ||||  | ||||  ||  | ||||  | ||||  ||  | ||||
Db        822 CCTGTCTGTCTGTCTGCCTGACTGCCTGTCTGTCTGCCTGACTGCCTGTCTGTCTGC 878


RESULT 57
BX173189/c
LOCUS       BX173189                551 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-174I13, genomic survey sequence.
ACCESSION   BX173189
VERSION     BX173189.1  GI:28004894
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 551)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 174I13. 174I13
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .551
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-174I13"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
BASE COUNT      230 a    162 c     51 g    108 t
ORIGIN

  Query Match             10.2%;  Score 102.4;  DB 29;  Length 551;
  Best Local Similarity   52.0%;  Pred. No. 3.1e-09;
  Matches  229;  Conservative    0;  Mismatches  211;  Indels    0;  Gaps    0;

Qy        359 GACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 418
              ||| |   | |||   |  ||   |   | |||   |   | ||| ||| ||| | | |
Db        547 GACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGATTGATTGATTGATT 488

Qy        419 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 478
              |  || ||| ||| | | | |  || ||| ||| | | | |  || ||| ||| | | |
Db        487 GATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATT 428

Qy        479 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 538
              |  || ||| ||| | | | |  || ||| ||| | | | |  || ||| ||| | | |
Db        427 GATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATT 368

Qy        539 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCT 598
              |  || |||    | | | |   ||||| ||| ||| | |  | |    | |   |
Db        367 GATTGATTGGCTGGCTGGCTGGCTGGTTGGTTGGTTGGTTAACTGGTTCGTT 308

Qy        599 GATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 658
              ||| |  || | |   || | || || |  || || | || || | |   || |
Db        307 GATTGAATGGTTGGTTGCATGGTTGGATGGTTAGTTGGTTGGTTGGTTGGTTGGTT 248

Qy        659 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCT 718
              | | || || | |   || || | || || | |   || |  | || || | |   || |
Db        247 GGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 188

Qy        719 GCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 778
              | |    || | |  | ||    | |||   |   | |||   | |||   |   | ||
Db        187 GGTTAATTGGTTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACT 128

Qy        779 GCTGGGCTGCTAGTAGGGCT 798
              | . | ||| | |  || ||
Db        127 GACTGACTGATGGACGGACT 108


RESULT 58
BX447644
LOCUS       BX447644                1200 bp    mRNA    linear   EST 22-MAY-2003
DEFINITION  BX447644 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
            clone CS0DG006YO06 5-PRIME, mRNA sequence.
ACCESSION   BX447644
VERSION     BX447644.1  GI:31025859
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1200)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0AG006BH03QP1.
FEATURES             Location/Qualifiers
     source          1. .1200
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DG006YO06"
                     /tissue_type="B CELLS (RAMOS CELL LINE)"
```

```
                     /cell_line="RAMOS CELL LINE"
                     /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6.vector.
                     Library was not normalized."
BASE COUNT       20 a    465 c    595 g     46 t     74 others
ORIGIN

  Query Match             10.2%;  Score 102.2;  DB 13;  Length 1200;
  Best Local Similarity   42.4%;  Pred. No. 3.4e-09;
  Matches  359;  Conservative   40;  Mismatches  447;  Indels    1;  Gaps    1;

Qy          4 CGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTC 63
              ||| ||     ||| |   |||  ||   | | | | | |||   | ||   |   |
Db        294 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 353

Qy         64 TCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGC 123
                |||||   || | | | |  ||  || || |    | ||   | || ||     | ||
Db        354 CGCGCGCGCGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 413

Qy        124 GCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGG 183
              |||  || | || |  |       | |  ||| |    |||| | |     | |  |  |
Db        414 GCGCGCGCGCGCGCGCGCGCGCGCGCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 473

Qy        184 GGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTG 243
              |   ||| | ||| ||  | | |    || |  ||   ||  ||  || ||| ||  |
Db        474 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 533

Qy        244 CTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGG 303
              |  ||    | |   |   ||  ||  || ||| ||  ||  ||  . | |   |   |
Db        534 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 593

Qy        304 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 363
              |  ||  ||    | |   |  ||  ||  || |||  ||  |:| ::    | |   |
Db        594 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGMTCRYGCGCGCGCGCG 653

Qy        364 TAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 423
                |   ||  ||  ||  ||  |   || :|:  ||  ||  ||  || ||| ||  |
Db        654 CGCGCGCGCGCGCGCGCGCGTGYGCGCGCGYGYGYGCGCGCGCGCGCGCGCGCGCGCGCG 713

Qy        424 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 483
              |  ||  || |||  ||  ||  ||  || |||  ||  ||  ||  || ||| ||  |
Db        714 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 773

Qy        484 CTTGCTTGCTCGCTGGCCTGCTTGCT-TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 542
              |  ||  || |||  || :||  |||  ||  |||  ||  ||  ||   ||  |||  ||
Db        774 CGCGCGCGCGCGCGCGCGYGCGCGCTGCGCGCGCGCGCGCGCGCGTGTGCGCGCGCGCGCGC 833

Qy        543 GCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATG 602
              ||  ||     | |   ||  |   ||  ||  ||     | |   |  |||  |   | |
Db        834 GCGCGCGCGCGCGCGCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGGGGGGGGG 893

Qy        603 GCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 662
              |   |   || ||  ||  |||  |   |  |||  |  |||  | | |  |||     |
Db        894 GGCGGGGGGTTGGTGTGGGGGGGGGGGGGGGGGGGGGGGGGKGGKTGGGTGGGGGTGGG 953

Qy        663 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTG 722
              |  ||   |  |||  |  ||| |   | |  ||| |   ||| :|      ||  |  ||
Db        954 GTGTGTGTGGTGGGGGGGGGGGGTTGTTGGGGGGTTTGGGGGGKGGGGTGGGGTGTTGTG 1013

Qy        723 GGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 782
              |     | |  |||    ||||  ::     :|  | :  | : ´  | : :  |
Db       1014 GTGGTGTGGGGGGGGGGGTGGGGGSYACAACVGMTTTMVCGTSGAGCAYKGRCACGTRCC 1073

Qy        783 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTT 842
              | :::: : :|::  :   |  :|| ::  : : |: | :| |  : :|  :|:  :
Db   .   1074 SCSYKYGCRSMGRSGCSAGCGCGKGCRVSRGSTSGTRKCGSCGGCGKSSCCCSCBCCSCG 1133

Qy        843 GCTAGTG 849
              |: ::::
Db       1134 GYGMSBR 1140
```

```
RESULT 59
BZ470200
LOCUS       BZ470200                705 bp    DNA     linear   GSS 13-DEC-2002
DEFINITION  BONQQ85TF BO_1.6_2_KB_tot Brassica oleracea genomic clone BONQQ85,
            genomic survey sequence.
ACCESSION   BZ470200
VERSION     BZ470200.1  GI:26767003
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 705)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BONQQ85TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .705
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BONQQ85"
                     /clone_lib="BO_1.6_2_KB_tot"
                     /note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared
                     total DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      112 a    203 c    213 g    177 t
ORIGIN

  Query Match             10.2%;  Score 101.8;  DB 29;  Length 705;
  Best Local Similarity   53.1%;  Pred. No. 4e-09;
  Matches  217;  Conservative    0;  Mismatches  192;  Indels    0;  Gaps    0;

Qy        223 GCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 282
              || |||  | ||||||    || ||||  |  |       || |||| | ||||||
Db         41 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 100

Qy        283 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 342
              || ||||  |  |       || |||| | |    |  ||  ||  | | |||   |
Db        101 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 160

Qy        343 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCT 402
              || |||| | |    |  ||   | | ||   |||  |   | ||||||
Db        161 GCGTGCTCGGTCGCAGCGTAGCGTGCTCGGTCGCAGCGTAGCGTGCTCGGTCGCTGCGTA 220

Qy        403 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 462
              || |||| | ||||||    || |||| | ||||||    || |||| | ||||||
Db        221 GCGTGCTCGGTCGCTGTGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 280

Qy        463 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 522
              || |||| | ||||||    || |||| | ||||||    || |||| | ||||||
Db        281 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 340

Qy        523 GCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGG 582
              || |||| | |||||     || |||| |  |       |  |||| | | |||  |
Db        341 GCGTGCTCGGTCGCTCCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTA 400

Qy        583 GACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAG 631
              |   |  ||   ||    | || || | |   |     | |  |||| |
Db        401 GCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTCCGTAGCGTGCTCG 449
```

```
RESULT 60
BZ179889
LOCUS       BZ179889                416 bp    DNA     linear   GSS 11-OCT-2002
DEFINITION  CH230-444A16.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-444A16, genomic survey sequence.
ACCESSION   BZ179889
VERSION     BZ179889.1  GI:23824087
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-444A16.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 444  row: A  column: 16
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .416
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-444A16"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT       57 a    118 c     66 g    175 t
ORIGIN

  Query Match             10.0%;  Score 100.6;  DB 28;  Length 416;
  Best Local Similarity   72.6%;  Pred. No. 6.5e-09;
  Matches  130;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        371 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 430
              ||| || |||| || ||  |||| | | || ||||||||||||| ||| || |||||||||
Db         64 GCTAGCCTGCTGGTGGGCTTGCTAGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 123

Qy        431 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              |||  ||  | ||||| ||||||  || || | ||||||| || |||  | ||||| |||
Db        124 GCTTTCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTT 183

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              |||  || || ||||||| || | ||| || ||||  ||| ||  ||  | | ||| ||
Db        184 GCTTCCTTGCTTGCTTGCCTGATGGCTAGCTTGCTATCTTCCTTCCTTCCTTCCTTCCT 242


RESULT 61
AI676338
```

```
LOCUS       AI676338                457 bp    mRNA    linear   EST 19-MAY-1999
DEFINITION  etmEST0084 EtH1 Eimeria tenella cDNA clone etmb031 5', mRNA
            sequence.
ACCESSION   AI676338
VERSION     AI676338.1  GI:4876818
KEYWORDS    EST.
SOURCE      Eimeria tenella
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   Wan,K.L., Chong,S.P., Ng,S.T., Tomley,F.M. and Jangi,M.S.
  TITLE     Survey of genes in Eimeria tenella merozoite
  JOURNAL   Unpublished
COMMENT     Contact: Wan KL
            Centre for Gene Analysis and Technology
            Universiti Kebangsaan Malaysia
            43600 UKM Bangi, Selangor DE, Malaysia
            Tel: 6 03 8292997
            Fax: 6 03 8293249
            Email: klwan@pkrisc.cc.ukm.my
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7
            Seq primer: SK.
FEATURES             Location/Qualifiers
     source          1. .457
                     /organism="Eimeria tenella"
                     /mol_type="mRNA"
                     /strain="Houghton"
                     /db_xref="taxon:5802"
                     /clone="etmb031"
                     /dev_stage="second generation merozoite"
                     /lab_host="XL1-Blue MRF'"
                     /clone_lib="EtH1"
                     /note="Vector: Lambda ZAPII; Site_1: EcoRI; Site_2: XhoI;
                     Second generation merozoites of E. tenella H were purified
                     by column  chromatography and mRNA extracted using a
                     FASTTRACK kit (Invitrogen).  cDNA was synthesised and a
                     Uni-ZAP XR library was constructed using  cDNA synthesis
                     kit, ZAP-cDNA synthesis kit and ZAP-cDNA Gigapack III
                     Gold cloning kit (Stratagene). The library was amplified
                     once through  E. coli XL1-Blue MRF'."
BASE COUNT       93 a    122 c    155 g     87 t
ORIGIN

  Query Match             10.0%;  Score 100.2;  DB 9;  Length 457;
  Best Local Similarity   61.9%;  Pred. No. 7.7e-09;
  Matches  159;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy        366 GGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 425
              || |  | ||   ||| | |  |     |||||   || |  |  |  |||||||  |||
Db        117 GGCCCACAGGGCAGCTGGGACTGGCTGGCGCTGAGGTGGTCTCGGGACCGCTGGCTAGCT 176

Qy        426 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 485
              ||| ||| ||| ||  |||||||| | | |||||    ||| ||| ||| |||||||  |||
Db        177 AGCTAGCTAGCTAGCTAGCTTGCTGGATAGCTGGATAGCTGGCTAGCTGGCTGGCTAGCT 236

Qy        486 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 545
              | | ||| ||| ||  ||| ||| ||||||||||  ||| ||| ||| ||| ||  |||
Db        237 GGGTCGCTGGCTAGCTAGCTGGCTAGCTCGCTGGCTAGCTAGCTGGCTAGCTAGCTGGCT 296

Qy        546 TGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCT 605
              |||    ||  ||  ||   ||| ||| ||||    | |   | ||  ||  | |||||
Db        297 AGCTAGCTGGCTAGCTGGCTGGCTAGCTAGCTAGCTAGCCACCAAGCTAGCTAGCTGGCT 356

Qy        606 TGATAGTAGGGCTGCTG 622
              |  ||| |   ||| |
Db        357 AGCGAGTTGCCTTGCCG 373


RESULT 62
BZ497497
```

```
LOCUS       BZ497497                748 bp    DNA     linear   GSS 16-DEC-2002
DEFINITION  BONKB78TR BO_1.6_2_KB_tot Brassica oleracea genomic clone BONKB78,
            genomic survey sequence.
ACCESSION   BZ497497
VERSION     BZ497497.1  GI:27010919
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 748)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BONKB78TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .748
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BONKB78"
                     /clone_lib="BO_1.6_2_KB_tot"
                     /note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared
                     total DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      153 a    210 c    218 g    167 t
ORIGIN
```

```
  Query Match           10.0%;  Score 100.2;  DB 29;  Length 748;
  Best Local Similarity   54.2%;  Pred. No. 7.8e-09;
  Matches  227;  Conservative    0;  Mismatches  188;  Indels    4;  Gaps    1;

Qy        311 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 370
              |||  |  |  |   || |||| | |||||     || ||||  |  | ||    |  |
Db        317 GCTCGGTAGCTACGTAGCATGCTCGGTCGCTACGTAGCGTGCTCGGTCGCCACGTAGCGT 376

Qy        371 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 430
              |||  |  | ||  |    |  |||||| | |||   | || ||||||  | |||   | || |
Db        377 GCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGT 436

Qy        431 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              |||||  |  |||    | || ||||||. | |||   | || ||||||  | |||   | || |
Db        437 GCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGT 496

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC----TGCTT 546
              |||||  |  |||    | || ||||||  |  |||   | || ||||||  |  |||     || |
Db        497 GCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGT 556

Qy        547 GCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTT 606
              |||  |  | ||    |  |||| | | || | |  |   |  ||   |||   | || |
Db        557 GCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGT 616

Qy        607 GATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 666
              | | |   | |   | |  |||| |   | |   | |  |||| |   | |   | | |
Db        617 GCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGT 676

Qy        667 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGC 725
              |||  |   | |   | |  |||| |   | | | | |  || | |   | |  |||||
Db        677 GCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGATCGGTCGCCACGTGGGC 735
```

```
RESULT 63
BX426372
```

```
LOCUS       BX426372                1115 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX426372 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
            clone CS0DG006YK02 3-PRIME, mRNA sequence.
ACCESSION   BX426372
VERSION     BX426372.1  GI:30772568
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1115)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0AG006BF01NP1.
FEATURES             Location/Qualifiers
     source          1. .1115
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DG006YK02"
                     /tissue_type="B CELLS (RAMOS CELL LINE)"
                     /cell_line="RAMOS CELL LINE"
                     /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the NOt I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
BASE COUNT        8 a    494 c    577 g      0 t     36 others
ORIGIN
```

```
  Query Match           10.0%;  Score 99.6;  DB 13;  Length 1115;
  Best Local Similarity   44.9%;  Pred. No. 1e-08;
  Matches  442;  Conservative    3;  Mismatches  537;  Indels    2;  Gaps    2;

Qy          4 CGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTC 63
              ||| ||     |||  |   |||  ||   | | | | | |||   | ||   |   |
Db        129 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 188

Qy         64 TCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGC 123
                |||||   || | | | |  || |  |  | ||||    ||| | |  | |||| |
Db        189 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 248

Qy        124 GCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGG 183
                |   | | || | |  || | |  || |    |||| | |     | | | |
Db        249 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 308

Qy        184 GGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTG 243
                |  ||| | ||| ||  | | |    || |  ||   ||  ||  || ||| ||  |
Db        309 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 368

Qy        244 CTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGG 303
              |  ||    | |   |  ||  || ||| ||  ||  ||  |  | |   |  |
Db        369 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 428

Qy        304 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 363
              |  ||  ||    | |   |  ||  || ||| ||  ||  ||  |  | |   |  |
Db        429 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 488

Qy        364 TAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 423
                |  ||  ||  ||  |   |  || ||| ||  ||  ||  || ||| ||  |
Db        489 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 548

Qy        424 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 483
                |  ||  || |||  ||  ||  ||  || |||  ||  ||  ||  || |||  ||  |
```

```
Db        549 CGCGCGCGCGCGCGMGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 608

Qy        484 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 543

Db        609 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 668

Qy        544 CTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGG 603

Db        669 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 728

Qy        604 CTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 663

Db        729 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 788

Qy        664 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGG 723

Db        789 CGCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC-GCGCGGGGGCGCGCGCG 847

Qy        724 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 783

Db        848 CGCGCGCGCGCGCGCGCGCGCGCGGGCGCGCGCGCGGGGGCGCGCGCGCGCGCGCCGG 907

Qy        784 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGC-TTGCTT 842

Db        908 GCAGCGMGCGCGACGCGCGGGCGGGGCGGGCGCCGGCGCGGCAGCGCGCGCGCGCGGCGG 967

Qy        843 GCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGGG 902

Db        968 GCGGSGGCCCCCGCGCGCGCGCGGGCGCGGCGCGCGGCGCGCGGGGCGCGCGGCGGCGGG 1027

Qy        903 CTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACCCGGGCC 962

Db       1028 CGGGCGGCGGGCGCCGCGCCGGCGGGGGCGGGGGGGGGGGGSGGCGCGCGCGSAGCGCGCG 1087

Qy        963 TCTCCCCCGCCCGGACCGCCGGGC 986

Db       1088 SCGCGCGCGACGGCGGCGCCGC 1111


RESULT 64
BX242384/c
LOCUS       BX242384                 702 bp    DNA     linear   GSS.29-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-282O18, genomic survey sequence.
ACCESSION   BX242384
VERSION     BX242384.1  GI:28164718
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 702)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 282O18. 282O18
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .702
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-282O18"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
BASE COUNT      221 a    282 c     44 g    155 t
ORIGIN

  Query Match             9.9%;  Score 99;  DB 29;  Length 702;
  Best Local Similarity   50.1%;  Pred. No. 1.3e-08;
  Matches  272;  Conservative    0;  Mismatches  270;  Indels    1;  Gaps    1;

Qy        318 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCT 377

Db        702 GGTTGGTTGGTTGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGAT 643

Qy        378 TGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 437

Db        642 GGATGGATGGATGGATGGATGGATGGATGGATGGATGGTTGGTTGGTTGGTTGGTT 583

Qy        438 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 497

Db        582 GGTTGGATGGATGGATGGATGGATGGTTGGTTGGATGGATGGATGGATGGATGGATGGAT 523

Qy        498 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGAC 557

Db        522 GGATGGATGGATGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGATGGAT 463

Qy        558 AGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGC 617

Db        462 AGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGAT-GGATGGA 404

Qy        618 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 677

Db        403 TGGATGGATGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGATGGATGGATGGA 344

Qy        678 TGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 737

Db        343 TGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGA 284

Qy        738 TCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 797

Db        283 TGGATGGATGGATGGATGGTTGGTGGGTTGGTTGGCGGATAACTTGGTTGGTTGGTTGGT 224

Qy        798 TGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 857

Db        223 TTGCTGGCTGGCTCACTAGCTGGCAGGCTGACTAATTGATTGATTGAGAGCAAGACAGAT 164

Qy        858 TGC 860

Db        163 GGC 161


RESULT 65
BX405071
LOCUS       BX405071                 1201 bp    mRNA    linear   EST 13-MAY-2003
DEFINITION  BX405071 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
            clone CS0DG006YG06 3-PRIME, mRNA sequence.
ACCESSION   BX405071
VERSION     BX405071.1  GI:30648111
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 10245.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0AG006BD03NP1&cluster=10245.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0AG006BD03NP1.
FEATURES             Location/Qualifiers
```

```
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DG006YG06"
                     /tissue_type="B CELLS (RAMOS CELL LINE)"
                     /cell_line="RAMOS CELL LINE"
                     /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
BASE COUNT       62 a    448 c    462 g      5 t    224 others
ORIGIN

  Query Match             9.8%;  Score 98.2;  DB 13;  Length 1201;
  Best Local Similarity   37.2%;  Pred. No. 1.8e-08;
  Matches  369;  Conservative   68;  Mismatches  554;  Indels    2;  Gaps    2;

Qy          3 CCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGT 62
              |    |       |      |  |    | |  |  | | |  | | |  |  |  |
Db         16 CNNNNCNCNNCNNCNCNNNNCNNCNCNNCNCNNCNNCNCNNCNCNNCNNCNCNNCNN 75

Qy         63 CTCCGCGCCGCCGGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGG 122
              | |  |  | || | | | |  || | |  | ||||   ||| |  |  ||||| |
Db         76 CNCNNCNNCNNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 135

Qy        123 CGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAG 182
                 |   |  | || | |  || | |  || |    ||||  | |     | |  |
Db        136 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 195

Qy        183 GGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCT 242
              | |  ||| | ||| || | | |  | || |  ||   ||  |   |  |||  |
Db        196 GCGCGCGCGCGCGCGCGCGCGCGCGAGCGAGCGCGAGAGCGAGNGCGNGCGCGNGNGC 255

Qy        243 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAG 302
              |   |     | |   |  ||  ||  ||  :|  ||  ||  ||    | |   |
Db        256 GNGNGNGNGCGCGCGCGCGCGCGCGCGCGCGAKCGCGCGCGCGCGCGCGCGCGCGCGC 315

Qy        303 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 362
              ||  ||  ||    | |   |  ||  ||  || |||  ||  ||  ||    | ||
Db        316 GCGCGCGCGCGCGCGCGCGCGAGCGCGCGCGAGCGCGCGCGCRCGCGCGCGCGCGCGAGC 375

Qy        363 GTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 422
              |   |   |   |   |   |   ||  |   |   |   |   |   |   ||  ||
Db        376 GNGCGNGNGNGNGAGNGNGNGNGNGNGCGNGAGNGNGCGNGNGAGAGAGNGCGCGCGC 435

Qy        423 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 482
              ||  ||  || |||  :|  ||  ||  ||  || |||  ||  ||  ||   | |||  ||
Db        436 GCGCGCGCGCGCGCACRCRCGCGCGCGCGCGCGCGCGCGCGCGCGCGCACGCGCGCGCGC 495

Qy        483 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 542
              ||  ||  || |||  ||  ||  ||  |: |||  ||  ||  ||  || |:|  ||
Db        496 GCGCGCGCGCGCGCGCGCGCGCGCGCGMGCGCGCGCGCGCGCGCGCGCGCRCGCGCGM 555

Qy        543 GCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATG 602
              ||  ||    | |   |   |   ||  ||  |:     |   :  |||: :|: :
Db        556 GCGCGCGCGCGCGCGCGCGCGCGCGCSCGCGCGMGCGCCCGCGMSMSCGCSASCSASCGC 615

Qy        603 GCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 662
              :|  |   |    :|  |    : ||  |   |   :|  | : ||  :   |   |: :
Db        616 SCACGCGCGSCCCSCCACSCCCSCGCGCGCGCGCGCSCGCGCSCGCGCSCGCGCGCGSCS 675

Qy        663 GGCTGCTAG-TAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCT 721
              :   || ||   | |:  : | || |: |     || |: | ||  :        |  |
Db        676 SCGCGCGAGVSCGCGMGCSAGCGCCGSCACGCACGCCCSCGCGCCCSCSMCCCACCSACC 735

Qy        722 GGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCT 781
              |: |: |    : | |||: ||  :     :::| :: ||:|        ||:| ::
Db        736 CCGSCGSCACMGCSACGCCTSCGCSCSCGCGCSSSCGSSSGGSCGCGGGCGCGGSCGSSG 795

Qy        782 GGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGC-TTGC 840
```

```
              : || :|     :||| || | :| :|      |   ||  || |:|   :  :|   ||
Db        796 SCGCGSCGCCGGGGGCGGCCGCSCGSCGCCGCCGCGCGCGCGCCGSCGCCSCGSCGSGGC 855

Qy        841 TTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAG 900
                :   |  |   |||   |  |:  :  |:|  |: | |:||  :| ||  | : |  |
Db        856 GCSGCTGGCGCCACGCGGCCGCGSCGSCGGSGAMGSSGCGSAGGGKGCAGSGCCRCGSGG 915

Qy        901 GGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACCCGGG 960
              :     ||| | |  ||| ::|  :|||  || |     | :||:|    |:  | ||||
Db        916 SCAAAGCGGGCGGGCGCGGSSGGCBGCGGGCGSGAMRGCGGSGARGCGGGGSAGCGCGGG 975

Qy        961 CCTCTCCCCCGCCCGGACCGCCGGGCACCCGGC 993
              |     |  :| :  | ::||:||||  | |::
Db        976 CGGGGGCGGSGGGSRCGGSSGCSGGGCBGCGGSS 1008
```

```
RESULT 66
AK089803/c
LOCUS       AK089803               2933 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus activated spleen cDNA, RIKEN full-length enriched
            library, clone:F830023K21 product:unclassifiable, full insert
            sequence.
ACCESSION   AK089803
VERSION     AK089803.1  GI:26105561
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
```

```
            TQyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2933)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
            Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
            Trust/MRC building Addenbrookes Hospital Cambridge) whose
            assistance we gratefully acknowledge.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .2933
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="NOD"
                     /db_xref="FANTOM_DB:F830023K21"
                     /db_xref="taxon:10090"
                     /clone="F830023K21"
                     /tissue_type="activated spleen"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
     misc_feature    1. .2933
                     /note="unclassifiable"
BASE COUNT      867 a    625 c    616 g    825 t
ORIGIN

  Query Match             9.8%;  Score 98.2;  DB 11;  Length 2933;
  Best Local Similarity   58.6%;  Pred. No. 1.9e-08;
  Matches  208;  Conservative    0;  Mismatches  143;  Indels    4;  Gaps    2;

Qy        334 TGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCT 393
              | ||   ||| ||  |  |    || | |     ||| |||   |  |   | | |||||
Db       1605 TACTTTATGGTCTAGTCTCAAGAGGTAAATCTCAGCTCCTGCTCTAGTGCCATGCCTGCT 1546

Qy        394 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 453
              ||  |||||| ||| || | ||| | ||| |||| ||  ||  |||||| ||| |||
Db       1545 TGCCTGCCTGCCTGCCTGTTGGCT-GTCTGTTTGCCTGCCTGCCTGCCTGCCTGCC 1487

Qy        454 CGCTGGCCTGCTTGCTTGCTCGCTG---GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 510
              ||  |||||| |||| |||  |||   |||||| ||| |||  ||| | |||  || ||
Db       1486 TGCCTGCCTGCCTGCTGGCTGTCTGTTTGCCTGCCTGCCTGCCTGCTTGGCTGTCTGTTT 1427

Qy        511 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTT 570
              ||  ||  |||||| ||  ||   | | |||||| |||      | | ||  | |||| |
Db       1426 GCCTGCCTGCCTGCCTGGCTGTCTGTTTGCCTGCCTGCCTGCCTGCCTGTTTGCCTGCCT 1367

Qy        571 GCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTA 630
              || |||   |  | | ||  |||||||| | || ||   |   | ||| | | ||||
Db       1366 GCCTGCCTGGCTGTCTGTTTACCTGCCTGCTTGCCTGCCTGGCTGTCTGTTTGCCTGCCT 1307

Qy        631 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGC 685
              |  | ||||   ||||  |   | |||||   | ||  |    || ||| |
Db       1306 GCCTGCCTGCCTTCCTGCCTGCCTGCCTGCTTCCATACTCCTCACTATGATGGTC 1252


RESULT 67
AK080786
LOCUS       AK080786               4346 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
            enriched library, clone:B130055L14 product:WASP family 1, full
            insert sequence.
ACCESSION   AK080786
VERSION     AK080786.1  GI:26099478
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
```

```
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 4346)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. Tomohiro Kono (Department of Animal
            Science, Tokyo University of Agriculture, 1737 Hunako Atsugi City,
            Kanagawa Prefecture, Japan) whose assistance we gratefully
            acknowledge.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .4346
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:B130055L14"
                     /db_xref="taxon:10090"
                     /clone="B130055L14"
                     /tissue_type="parthenogenote"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="9.5 days embryo"
     misc_feature    1. .4346
                     /note="WASP family 1 (MGD|MGI:1890563, GB|NM_031877,
                     evidence: BLASTN, 99%, match=394)"
BASE COUNT     1075 a    854 c   1012 g   1405 t
ORIGIN

  Query Match             9.8%;  Score 98;  DB 11;  Length 4346;
  Best Local Similarity   56.7%;  Pred. No. 2.1e-08;
  Matches  220;  Conservative    0;  Mismatches  165;  Indels    3;  Gaps    2;

Qy      314 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 373
            ||  ||    | |||   |  ||| |||||  | | | ||   ||   || ||   | |
Db      868 GACTGGTTGATTGGCTGCCCGGCTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 927

Qy      374 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT--GCTTG 431
                || | | | || |    ||| |||  |   || | | ||| ||| ||||| ||  |   |
Db      928 GGTTGGTTGGTTGTCTGGCTGGCTACCTGCCTGGTTGGCTGGCTAGCCTGGTTCGGTCGG 987

Qy      432 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 491
            || | |||   |   | | ||| ||||||  ||| ||| ||| ||| ||||||   || ||| |
Db      988 CTGGGTGGTTGGTCGGTTGGCTGGCTGGCTAGCTGGCTGGCTGGCTGGCTGACTGGCTGG 1047

Qy      492 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGA 551
            || ||||||  ||| ||| ||| | |||   | | | | | | ||||||  ||| ||||
Db      1048 CTGGCTGGCTGGCTGGCTGGCTGGTTGGATGGTTGGTTGGGTGGCTGGCTGGCTGGCTGG 1107

Qy      552 GGGGACAGTAGGGCTGCTT-GCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 610
              ||   |   || || || ||||| |  || | | |   | ||| ||| | | |||
Db      1108 CTGGCTGGGCTGGTTGGTTGGCTTGGTTGGGTGGCTGGCTGGCTGGCTGTTTGGTTGGGT 1167

Qy      611 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 670
            |   |||||   |||||   |   ||   ||||   ||| |   |     ||    |||
Db      1168 GACCGGCTGGCTGGCTGACTGATTGGTGTCTGGCAGGCTGGCTAGTTGTGTGCTTGGCTG 1227

Qy      671 GTAGGGCTGCTGGGCTGCTAGTAGGGCT 698
            |  |  | |||||   ||| || ||| |
Db      1228 GCTGACCAGCTGGCTGGCTGGTTGGGTT 1255


RESULT 68
BZ216441/c
LOCUS       BZ216441                592 bp    DNA     linear   GSS 11-OCT-2002
DEFINITION  CH230-323I7.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-323I7, genomic survey sequence.
ACCESSION   BZ216441
VERSION     BZ216441.1  GI:23874799
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-323I7.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 323  row: I  column: 7
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-323I7"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
```

```
BASE COUNT    201 a    177 c    104 g    110 t
ORIGIN

  Query Match             9.8%;  Score 97.8;  DB 29;  Length 592;
  Best Local Similarity   60.2%;  Pred. No. 2.1e-08;
  Matches  162;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy        265 CTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGT 324
              |||||   |     |  |||| | | |  |||  |  ||  |  | | | |  |||
Db        394 CTGCTGCTTTTGCTGTTTGCTGGATTACTGGATTGCTGGATTAATGGATTACTGGATTAC 335

Qy        325 AGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGT 384
               ||  |||| | | |||||  |||| | |    |||  |  ||  |||||| ||||| |
Db        334 TGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGA 275

Qy        385 AGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 444
                |   | |||||||  | || | || || | |  |  | ||||| | | |||||  |||
Db        274 TCGCTGGATCGCTGGATTACTGGATTACTGGATTACTGGATTGCTGGATTGCTGGATTGC 215

Qy        445 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 504
              | | ||||| | | ||  | ||||| | | ||||| ||||  | ||||| | | ||  |
Db        214 TGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGA 155

Qy        505 TTGCTTGCTCGCTGGCCTGCTTGCTTGCT 533
              ||||| | | |||||  |||| | |||||
Db        154 TTGCTGGATTGCTGGATTGCTGGATTGCT 126


RESULT 69
CD347163/c
LOCUS       CD347163                 566 bp    mRNA    linear   EST 29-MAY-2003
DEFINITION  EtESTeflc03.yl Eimeria tenella M5-6 Excised cDNA Eimeria tenella
            cDNA 5' similar to TR:Q25062 Q25062 PRECURSOR PROTEIN FOR RFAMIDE.
            ;, mRNA sequence.
ACCESSION   CD347163
VERSION     CD347163.1  GI:31138784
KEYWORDS    EST.
SOURCE      Eimeria tenella
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 566)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis,. MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .566
                     /organism="Eimeria tenella"
                     /mol_type="mRNA"
                     /db_xref="taxon:5802"
                     /dev_stage="LS18"
                     /lab_host="E.coli DH10B (GeneHog, Invitrogen,Inc)"
                     /clone_lib="Eimeria tenella M5-6 Excised cDNA"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Sporozoites were obtained from in vitro sporulated
                     and excysted oocysts of E.tenella grown in chickens. cDNA
                     was synthesized from poly mRNA using an oligo-dT primer
                     containing a XhoI site. Following second strand synthesis,
                     EcoRI adapters were ligated to the cDNA and products were
                     size-selected on sephacryl S500. The cDNA were ligated to
                     EcoRI/XhoI prepared lambda ZapII (Stratagene). The primary
                     library was mass excised using ExAssist helper phage
                     (Stratagene). The phagemids were preciptated with PEG 8000
                     , extracted with phenol/chloroform and electroporated into
                     DH10B cells. The library may contain a small percentage of
                     host or bacterial contaminants. Library materials provided
                     by: Paul Liberator, Merck Research Labs Library
                     constructed by: Paul Liberator, Merck Research Labs."
BASE COUNT      174 a    186 c    134 g     72 t
ORIGIN

  Query Match             9.7%;  Score 97.4;  DB 14;  Length 566;
  Best Local Similarity   53.6%;  Pred. No. 2.5e-08;
  Matches  203;  Conservative    0;  Mismatches  176;  Indels    0;  Gaps    0;

Qy        199 GCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACA 258
              ||| |   |    | |||  |   ||  |  |||||||||   | | ||   | ||
Db        552 GCTAGATGGCTCGCTGGTTGGTACGCTAGATGGCTCGCTGGTTGGTTCGCCAGGTGGCTC 493

Qy        259 GTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGG 318
              |   |     | ||| | ||||  | | | ||    | |||  ||| |||| | ||   |
Db        492 GATAGTTGTTTCGCTAGATGGCTCGATAGTTGGTTCGCCAGCTGGCTTGCTGGGTGTTTG 433

Qy        319 GACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTT 378
              |  ||  |    |||  || | | ||  | |   |     |   ||  |   | |||||
Db        432 GCTAGGTGAATCGCTAACTGGTTCGCTGGGTGATTCGCTAGTTGGTTCGATAGGTGGCTC 373

Qy        379 GCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 438
              |||||| ||   |||   |||   ||| | | | |||||  |   | | ||| ||| |||
Db        372 GCTAGTTGGTTCGCTAATTGGTTCGCTAGTTGGTTCGCTAGTTGGTTCGCTAGCTGGCTC 313

Qy        439 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 498
              ||  | ||| | ||| | |||   | | |||  ||||||||     | ||| | |   |
Db        312 GCTGGGTTGTTCGCTAGTTGGGTGGGTAGCTGACTCGCTGGGTGTTTGGCTAGGTGAGTC 253

Qy        499 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 558
              ||  ||| |||  ||  ||| |  ||| ||  |  |||| ||| ||| ||||  |  | |
Db        252 GCTAGCTGGCTCTCTATCTGTCTCGCTAGCCGGTCCGCTAGCCGGCTCGCTGGTGTAAAA 193

Qy        559 GTAGGGCTGCTTGCTTGCT 577
                |     || ||| |||
Db        192 ACTGTCTGACTCGCTAGCT 174


RESULT 70
BH704622/c
LOCUS       BH704622                 394 bp    DNA     linear   GSS 20-FEB-2002
DEFINITION  BOMBE06TR BO_2_3_KB Brassica oleracea genomic clone BOMBE06,
            genomic survey sequence.
ACCESSION   BH704622
VERSION     BH704622.1  GI:18786296
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 394)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOMBE06TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
```

```
FEATURES             Location/Qualifiers
     source          1. .394
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOMBE06"
                     /clone_lib="BO_2_3_KB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       96 a    137 c    118 g     43 t
ORIGIN

  Query Match             9.7%;  Score 97.2;  DB 28;  Length 394;
  Best Local Similarity   55.0%;  Pred. No. 2.7e-08;
  Matches  220;  Conservative    0;  Mismatches  168;  Indels   12;  Gaps    1;

Qy        391 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 450
              | |||||     || |||| | |||||     || |||| | |||||     || ||||
Db        389 GGTCGCTACGTAGCGTGCTCGGTCGCTACGAAGCGTGCTCGGTCGCTACGAAGCGTGCTC 330

Qy        451 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 510
              | |||||     || |||| | |||||     || || | | |||||     || ||||
Db        329 GGTCGCTACGAAGCGTGCTCGGTCGCTACGTAGCGTGATCGGTCGCTACGTAGCGTGCTC 270

Qy        511 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTT 570
              | |||||     || |||| | |||||     || ||||  |  |  |     |  ||||
Db        269 GGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTC 210

Qy        571 GCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTA 630
              | | |||| |             ||| || | | | | ||    |  |||| ||  ||||
Db        209 GGTCGCTACG------------TTGCGTGCTCGGTCGCTACGTTGCGTGCTCGGTCGCTA 162

Qy        631 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 690
              |  |||| ||  ||||    |  |||| ||  ||||  | |  |||| ||  ||||
Db        161 CGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGAAGCGTGCTCGGTCGCTA 102

Qy        691 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTA 750
              |  | || ||  ||||    |  |||| ||  ||||  | |  | || ||  ||||
Db        101 CGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGAAGCGTGCTCGGTCGCTA 42

Qy        751 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 790
              |  |||| ||  ||||    |  |||| ||  ||||
Db         41 CGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTA 2


RESULT 71
BZ457666/c
LOCUS       BZ457666                787 bp    DNA     linear   GSS 13-DEC-2002
DEFINITION  BONRC33TF BO_1.6_2_KB_tot Brassica oleracea genomic clone BONRC33,
            genomic survey sequence.
ACCESSION   BZ457666
VERSION     BZ457666.1  GI:26737270
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 787)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BONRC33TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
```

```
FEATURES             Location/Qualifiers
     source          1. .787
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BONRC33"
                     /clone_lib="BO_1.6_2_KB_tot"
                     /note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared
                     total DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      169 a    246 c    238 g    134 t
ORIGIN

  Query Match             9.7%;  Score 97;  DB 29;  Length 787;
  Best Local Similarity   50.1%;  Pred. No. 3e-08;
  Matches  270;  Conservative    0;  Mismatches  265;  Indels    4;  Gaps    1;

Qy        331 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCT 390
              ||||  | ||| ||  | | |||  |  |   |   ||  |||   | ||  |   ||
Db        573 GCTTCATAGCTTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTC 514

Qy        391 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 450
              |||   | || |||| | | |||   | || |||  | | |||   | || |||  | |
Db        513 GCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTC 454

Qy        451 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 510
              |||   | || |||  | | |||   | || |||  | | |||   | || |||  | |
Db        453 GCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTC 394

Qy        511 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTT 570
              |||   | || |||  | | |||   | || |||  | |         ||   | | | |
Db        393 GCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCTCTACGTAGCGTGCCCGGTC 334

Qy        571 GCTTGCTAAGGGGACGGTACGC----CTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCT 626
              |||   ||  | | | |  |||      || ||   | | | ||    |  |||  ||
Db        333 GCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTC 274

Qy        627 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 686
              ||||    |  |||  ||  ||||    |  |||  ||  ||||    |  |||  ||
Db        273 GCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTC 214

Qy        687 GCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCT 746
              ||||    |  | |  ||  ||||    |  |||  ||  ||||    |  | |  ||
Db        213 GCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTC 154

Qy        747 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 806
              ||||    |  |||  ||  |||     |  |||  ||  ||||    |  |||  ||
Db        153 GCTACGTAGCGTGCCCGGTCACTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTC 94

Qy        807 GCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCT 865
              ||||    |  |||| | | |||   | || |||  | | |    |  || |||| | |
Db         93 GCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGT 35


RESULT 72
AZ437267/c
LOCUS       AZ437267                477 bp    DNA     linear   GSS 03-OCT-2000
DEFINITION  1M0225L23F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0225L23 F, genomic survey sequence.
ACCESSION   AZ437267
VERSION     AZ437267.1  GI:10561280
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 477)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
```

```
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0225  row: L  column: 23
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 477.
FEATURES             Location/Qualifiers
     source          1. .477
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0225L23"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      181 a     91 c    113 g     92 t
ORIGIN

  Query Match             9.7%;  Score 96.8;  DB 28;  Length 477;
  Best Local Similarity   74.4%;  Pred. No. 3.2e-08;
  Matches  122;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy        379 GCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 438
              | || || |  ||   |   | |||  || ||| | | | |  || ||||||||| |||
Db        164 GATAATACGTTTGTCTGTCTGTCTGTCTGTTTGTTTGTTTGTTTGTTTGCTTGCTTGCTT 105

Qy        439 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 498
              ||||||||||||||| ||  || ||||||||||  ||| || |||||| |||| |||
Db        104 GCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTT 45

Qy        499 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 542
              || ||| |||||||| ||| ||||||||||||||| ||  ||||
Db         44 GCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCCT 1


RESULT 73
BX405072
LOCUS       BX405072               1201 bp    mRNA    linear   EST 13-MAY-2003
DEFINITION  BX405072 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
            clone CS0DG006YG06 5-PRIME, mRNA sequence.
ACCESSION   BX405072
VERSION     BX405072.1  GI:30652139
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 10245.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0AG006BD03QP1&cluster=10245.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0AG006BD03QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DG006YG06"
                     /tissue_type="B CELLS (RAMOS CELL LINE)"
                     /cell_line="RAMOS CELL LINE"
                     /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
BASE COUNT      32 a    312 c    468 g    138 t    251 others
ORIGIN

  Query Match             9.7%;  Score 96.6;  DB 13;  Length 1201;
  Best Local Similarity   44.0%;  Pred. No. 3.6e-08;
  Matches  261;  Conservative   29;  Mismatches  303;  Indels    0;  Gaps    0;

Qy          4 CGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTC 63
              :|: ||     |:|  : : :|   || : | | | : | |||   | :|   |   |:
Db        256 YGYGCGCGCGCGYGCGYGYGYGTGYGCGYGCGCGCGYGCGCGCGCGCGYGYGCGYGTGYG 315

Qy         64 TCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGC 123
                | |:||   || | | | |  || |     | |:|  |  ||| |     ||||| |
Db        316 TGTGYGCGCGCGTGTGYGTGTGCGCGTGTGTGTGYGTGTGTGCGCGTGTGYGCGCGCGTG 375

Qy        124 GCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGG 183
                  |   | | || | |  || | |  || |    |||| | |     | |  | |
Db        376 TGTGCGCGYGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 435

Qy        184 GGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTG 243
                |   ||| | ||| ||  | | |    || |  ||   ||  ||  || |||  ||  |
Db        436 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 495

Qy        244 CTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGG 303
                | ||    | |   |   ||  ||  || |||  ||  ||  ||    | |   |  |
Db        496 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 555

Qy        304 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 363
                | ||  ||    | |   |   ||  ||  || |||  ||  ||  ||    | |   |
Db        556 CGCGCGCGCGCGCGCGCGYGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGYGCGCGCG 615

Qy        364 TAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 423
                |  ||  ||  ||  |   |   ||  ||  ||  ||  ||  ||  ||  |:  |
Db        616 CGCGCGCGCGMGCGCGCGCGCGYGCGCGCGCGCGCGCGCGCGCGCGCGCGCGYGCG 675

Qy        424 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 483
              | || :|| ||:: ||  ||::||  ||:|||  ||  ||  ||  :| |||  ||  |
Db        676 CGCGCGYGCGCGYKCGCGCGCKYGCGCGCKCGCGTGCGCGCGCGCGCKCGCGCGCGCGCG 735

Qy        484 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 543
              : |||  |: || : ::  ||: ||  || |||  ||  |:  ||  || |||  ||  |
```

```
Db        736 YGTGCGCGYGCGTKTKMGMGCKCGCGCGCGCGCGCGCGCGMGCGCGCGCGCGCGCGCGCG 795

Qy        544 CTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGC 596
              |  ||    | |   ||  |  :||  |||||| :  | |   |  |||  ||
Db        796 CGCGCGCGTGCGCGCGTKCGTTYGCGCGCTTGCGMGTGCGCGCGCGCGCGCGC 848


RESULT 74
CA054386
LOCUS       CA054386               580 bp    mRNA    linear   EST 04-MAR-2003
DEFINITION  ssalrgb517217 mixed_tissue Salmo salar cDNA, mRNA sequence.
ACCESSION   CA054386
VERSION     CA054386.1  GI:24384629
KEYWORDS    EST.
SOURCE      Salmo salar (Atlantic salmon)
  ORGANISM  Salmo salar
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
REFERENCE   1  (bases 1 to 580)
  AUTHORS   GRASP Consortium, Davidson,W.S., Koop,B.F. and
            http://web.uvic.ca/cbr/grasp.
  TITLE     A survey of Salmo salar transcripts from high complexity cDNA
            libraries
  JOURNAL   Unpublished
COMMENT     Contact: Koop BF
            Centre for Biomedical Research
            University of Victoria
            PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
            Tel: 250 472 4067
            Fax: 250 472 4075
            Email: bkoop@uvic.ca
            Genome Sciences Centre, BC Cancer Agency
            cDNA preparation, sequencing and bioinformatics:
            Y Butterfield, R Kirkpatrick, J Asano, N Girn, R Guin, D Lee,
            S Lee, T Olson, P Pandoh, A Prahbu, D Smailus, L Spence, J Stott,
            S Taylor, G Yang, J Schein, S Jones and M Marra.
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .580
                     /organism="Salmo salar"
                     /mol_type="mRNA"
                     /strain="McConnell"
                     /db_xref="taxon:8030"
                     /clone_lib="mixed_tissue"
                     /note="Vector: pCMVsport6; Library Creator:  Research
                     Genetics ; Atlantic salmon tissue contributors: Carlo
                     Biagi, Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.),
                     Simon Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery
                     (Crofton, B.C.), Rachel Roper (University of Victoria)"
BASE COUNT      148 a    125 c    156 g    151 t
ORIGIN

  Query Match             9.6%;  Score 96.2;  DB 14;  Length 580;
  Best Local Similarity   64.7%;  Pred. No. 4.2e-08;
  Matches  143;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        331 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCT 390
              ||| ||| |||| |  ||| |||||  ||   | |||   ||||||| ||| |  |
Db        313 GCTGGCTGGCTGACTGGCTGGCTGACTGGCTGGCAGGCTGGCTGGCTGGCTGACTA 372

Qy        391 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 450
              || |||| |  ||| |||  ||  |||||  ||  ||| ||| ||||||   || |||
Db        373 ACTGGCTGACTGGCTGGCTGACTGACTGGCTGGCAGGCTGGCTGGCTGGCTGACTGGCTG 432

Qy        451 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 510
              ||  |||||  ||| |||  ||  ||| |   || ||| ||| ||||||   || |||
Db        433 ACTGACTGGCTGGCTGGCTGACTAACTGACTGACTGGCTGGCTGGCTGGCTGACTGGCTG 492

Qy        511 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGA 551
              |||  ||||   ||| ||| |||  |||||  ||| |||||
Db        493 GCTGACTGGTTGGCTGGCTGGCTGACTGGCTGGCTGGCTGA 533
```

```
RESULT 75
CNS05B34
LOCUS       CNS05B34               1094 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T3 end of clone
            020G19 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL329305
VERSION     AL329305.1  GI:8222927
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1094)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1094
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="020G19"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA020AD10A1~end : T3"
BASE COUNT      142 a    321 c    296 g    317 t     18 others
ORIGIN

  Query Match             9.6%;  Score 96.2;  DB 29;  Length 1094;
  Best Local Similarity   54.6%;  Pred. No. 4.2e-08;
  Matches  191;  Conservative    0;  Mismatches  159;  Indels    0;  Gaps    0;

Qy        499 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 558
              |||||  ||  |||  ||  | |||  ||| |||  ||  | |||  ||       | |
Db        144 GCCTGTCTGTCTGCCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGTCTGTCTGTCT 203

Qy        559 GTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCT 618
              ||  | |||| ||  ||       | | ||  | |||||||   |  ||   ||  | ||
Db        204 GTCTGTCTGCCTGTCTGTCTGTCTGTCTGTCTGTCTGCCTGTCTGTCTGTCTGTCTGTCT 263

Qy        619 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 678
              ||  | |||   |   | ||||  | ||||  ||  | |||   | ||||  ||  | ||
Db        264 GCCTGTCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGTCTGTCT 323

Qy        679 GCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 738
              ||  | ||||  ||  | ||    | ||||  ||  | ||||  | ||||  ||  | ||
```

```
Db        324. GCCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGTCT 383

Qy        739 CCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 798
              | ||||  |   | ||||  | |||   ||  | ||||  | |||   ||  | ||
Db        384 GTCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTGTCTGCCTGCCTGCCTGTCNGTCTGTCT 443

Qy        799 GCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGT 848
              ||  | ||||  ||  | ||| |||   |  ||| ||| ||  |||  ||
Db        444 GCCTGCCTGCCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGTCTGCCTGT 493


RESULT 76
BG561076
LOCUS       BG561076                393 bp    mRNA    linear   EST 10-APR-2001
DEFINITION  EtESTed91c07.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
            tenella cDNA 5', mRNA sequence.
ACCESSION   BG561076
VERSION     BG561076.1  GI:13590074
KEYWORDS    EST.
SOURCE      Eimeria tenella
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 393)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 282.
FEATURES             Location/Qualifiers
     source          1. .393
                     /organism="Eimeria tenella"
                     /mol_type="mRNA"
                     /strain="LS18"
                     /db_xref="taxon:5802"
                     /dev_stage="Sporozoite stage"
                     /lab_host="SOLR"
                     /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                     ; Sporozoites were obtained from in vitro sporulated and
                     excysted oocysts of E.tenella grown in chickens. cDNA was
                     synthesized from poly mRNA using an oligo-dT primer
                     containing a XhoI site. Following second strand synthesis,
                     EcoRI adapters were ligated to the cDNA and products were
                     size-selected on sephacryl S500. The cDNA were ligated to
                     EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were
                     converted to phagemids by mass excision using ExAssist
                     helper phage and E.coli SOLR cell (Stratagene). Clones
                     were selected by negative hybridization against a pool of
                     overrepresented ESTs (N>=10,from 1682 previous reads).
                     Insert sizes range from 1.2-2.9Kb. The library may contain
                     a small percentage of host or bacterial contaminents."
BASE COUNT       81 a    119 c    120 g     73 t
ORIGIN

  Query Match             9.6%;  Score 95.6;  DB 10;  Length 393;
  Best Local Similarity   60.3%;  Pred. No. 5.3e-08;
  Matches  158;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy        335 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTC 394
              ||  ||| ||  ||| ||||    |  ||   |  |||   || |||||  ||   |||
Db        107 GCCAGCTAGCTAGCTAGCTGGCTAGCTAGCTAGCTTGCCATCTAGCTAGCTGGTTGGCTA 166

Qy        395 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 454
              ||| ||  ||| ||| ||  ||   || |||  || ||| ||| ||| ||  ||| |||
Db        167 GCTAGCTGGCTAGCTAGCCAGCGAACCGGCTGCCTGGCTGGCTAGCCAGCGAGCTAGCTG 226

Qy        455 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 514
              | ||||   || | | ||| ||| ||| ||  ||| ||| | ||||  ||| | | |||
Db        227 GGTGGCTAACTGGATGGCTGGCTAGCCAGCGAGCTAGCTGGGTGGCTAGCTGGATGGCTA 286

Qy        515 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 574
              ||||||  ||    | ||| ||  ||  ||| ||| ||  | ||   | || ||| ||
Db        287 GCTGGCTAGCCACATAGCTAGCCAGCTGGCTGGCTCAGCAGCCACCTGAGCAGCTAGCCA 346

Qy        575 GCTAAGGGGACGGTACGCCTGC 596
              ||||  |||   |   ||  ||
Db        347 GCTACTGGGGGAGGCAGCGCGC 368


RESULT 77
AZ857141/c
LOCUS       AZ857141                421 bp    DNA     linear   GSS 21-FEB-2001
DEFINITION  2M01610Z3R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M01610Z3 R, genomic survey sequence.
ACCESSION   AZ857141
VERSION     AZ857141.1  GI:13049041
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 421)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0161  row: O  column: 23
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 421.
FEATURES             Location/Qualifiers
     source          1. .421
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M01610Z3"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
```

10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

```
BASE COUNT      150 a     88 c    115 g     68 t
ORIGIN

  Query Match             9.6%;  Score 95.6;  DB 28;  Length 421;
  Best Local Similarity   73.5%;  Pred. No. 5.3e-08;
  Matches  122;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy        377 TTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 436
              ||| |  |   | |  |    ||  || |||||||||| ||  || ||| |||||||  ||
Db        194 TTGGTTCTGTTGTTATTGTTGGGTGTGTTTGCTTGCTTGCCTGCTTGCCTGCTTGCCTGC 135

Qy        437 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
              | ||||||||||| |||| ||  || ||| |||||||  ||| ||||||||||| |||| ||
Db        134 TTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGC 75

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 542
                || ||| ||||||||| ||| |||||||||||||||  ||  ||||
Db         74 CTGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCCT 29


RESULT 78
BX171642/c
LOCUS       BX171642                728 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-127F16, genomic survey sequence.
ACCESSION   BX171642
VERSION     BX171642.1  GI:28003347
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 728)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 127F16. 127F16
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .728
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-127F16"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
BASE COUNT      337 a    185 c     78 g    128 t
ORIGIN

  Query Match             9.5%;  Score 95;  DB 29;  Length 728;
  Best Local Similarity   56.0%;  Pred. No. 6.9e-08;
  Matches  201;  Conservative    0;  Mismatches  155;  Indels    3;  Gaps    1;

Qy        402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461
              | |||||||  | ||| || || ||| | ||| | | |    | || ||| |   | ||
Db        724 TACTTGCTTAGTTGCTTGCTTGGTTGGTCGCTTGGTTGGTCACATGGTTGGTTTGTTGCT 665

Qy        462 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 521
              |||||| ||| ||||| |  || |  |||| |   | || || ||  ||| |||||
Db        664 TGCTTGTTTGGTCGCTTGGGTGGTCACTTGGTTTGTTGCTTGGTTAATTGGTCGCTTGGT 605

Qy        522 TGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGG 581
              || ||| ||| | | | |  || ||| |  |  |   |   || |||||||||| ||  |
Db        604 TGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGCTTGCTTCCTTGGT 545

Qy        582 GGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCT 641
              |   |  |   ||  || | ||||| | |  |   || |  | |||   | |   ||||
Db        544 TGGTTGGTCTGTTGTTTGGTCGCTTGTTTGATGACTTGGTTAGTTGC---TTGCTTTGCT 488

Qy        642 GGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCT 701
              | ||||    | | |||| |  |||| ||  || || | || || | |   |  | |
Db        487 TTGTTGCTTTGTGTGTTGCTTGCTTGCTTGTTTGGTTGGTTGGTTGGTTGCTTGCTTAGT 428

Qy        702 GGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 760
              || |||| |   |  || | |  | || |   || | || |  || | |     ||||
Db        427 TGGTTGCTTGCTTGTTTGGTTGCTTACTTGCTTGGTTACTTGCTTGGTTGGTAATCTGC 369


RESULT 79
BH115707/c
LOCUS       BH115707                310 bp    DNA     linear   GSS 19-JUL-2001
DEFINITION  RPCI-24-298L24.TJ RPCI-24 Mus musculus genomic clone RPCI-24-298L24
            , genomic survey sequence.
ACCESSION   BH115707
VERSION     BH115707.1  GI:14955201
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 310)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished
COMMENT     Other_GSSs: RPCI-24-298L24.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 298  row: L  column: 24
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .310
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-298L24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /clone_lib="RPCI-24"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT       97 a     75 c    134 g      4 t
ORIGIN

  Query Match             9.5%;  Score 94.8;  DB 28;  Length 310;
  Best Local Similarity   59.8%;  Pred. No. 7.3e-08;
  Matches  159;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;
```

```
Qy        263 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACA 322
              |||||| |||  ||  |||||| |||      | | |   |||||| ||||     | |
Db        301 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGTCTGCCT 242

Qy        323 GTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTA 382
              ||  |||||| |||   ||| |   ||  |||   |     | | |||   || |||
Db        241 GTCTGCCTGCCTGCCTCCTGTCTGCCTGTCTGCCCGCCTGATCTGCCTGTCTGCCTGCCT 182

Qy        383 GTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 442
              |   | |||   ||  |||||| ||| |||  ||  |||||| ||| |||  ||  ||||
Db        181 GCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 122

Qy        443 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 502
              || ||| |||  ||  | |||| ||  |||  ||  |||||| ||  |||  |   ||||
Db        121 GCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCT 62

Qy        503 GCTTGCTTGCTCGCTGGCCTGCTTGC 528
              || ||  ||  |  |||||  ||| |
Db         61 GCCTGTCTGACCTTTGGCCCCCTTTC 36


RESULT 80
AI798117
LOCUS       AI798117                615 bp    mRNA    linear   EST 20-DEC-1999
DEFINITION  wh81d08.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2387151 3'
            similar to contains Alu repetitive element;contains element TAR1
            repetitive element ;, mRNA sequence.
ACCESSION   AI798117
VERSION     AI798117.1  GI:5363589
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 615)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
            M.D., Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1082   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 458.
FEATURES             Location/Qualifiers
     source          1. .615
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2387151"
                     /tissue_type="B-cell, chronic lymphotic leukemia"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_CLL1"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTGCTTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT       77 a    229 c    157 g    152 t
ORIGIN
```

```
  Query Match             9.4%;  Score 94.2;  DB 9;  Length 615;
  Best Local Similarity   54.8%;  Pred. No. 9.6e-08;
  Matches  210;  Conservative    0;  Mismatches  168;  Indels    5;  Gaps    1;

Qy        386 GGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 445
              || ||     |||| || || || | |   | || || || || | |   | || || ||
Db         79 GGCCTCGCATCTGGGCTCCTGGCCTCCCACCCGGGCTCCTGGCCTCCCACCCGGGCTCCT 138

Qy        446 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 505
                | |  || || || || |  | |   |||| || || || | |     || || ||
Db        139 GAGCTCCCAGCCGGGCTCCTGGGCTTCCATCTGGGCTCCTGGCCTCCCATGAGGGCTCCT 198

Qy        506 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAG-----GGGACAGT 560
                || |  .  |||| || || || |     |||| || || ||         |||    |
Db        199 GGCCTCACATCTGGGCTCCTGGCCTCACATCTGGGCTCCTGGCCTCACATCTGGGCTCCT 258

Qy        561 AGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGC 620
               |  || |   || |      || |    ||| | ||  |||| | | |   ||||| |
Db        259 TGCCCTCCCACCTGGGCTCCTGGGCTCCCAGCCAGGCTCCTGGCCTCACATCTGGGCTCC 318

Qy        621 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 680
              ||| || | |    |||| |||| ||   |   |||| |||| ||   |   ||||| |
Db        319 TGGCCTCCCATCCAGGCTCCTGGCCTTGAATCTGGGCCCCTGGCCTCACATCTGGGCTCC 378

Qy        681 TGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCC 740
              ||| || | |  |||||||||| || | |   ||| | |||| |||  |    |||||
Db        379 TGGCCTCCCATCAGGGCTCCTGGACTCCCATCTGGGATCCTGGCCTGGAATCCAGGCTCT 438

Qy        741 TGGGCTGCTAGTAGGGCTGCTGG 763
              ||| || | |    |||| || |
Db        439 TGGCCTCCCATCCAGGCTCCTAG 461


RESULT 81
AZ282406
LOCUS       AZ282406                656 bp    DNA     linear   GSS 27-JUL-2000
DEFINITION  RPCI-23-138J12.TJ RPCI-23 Mus musculus genomic clone RPCI-23-138J12
            , genomic survey sequence.
ACCESSION   AZ282406
VERSION     AZ282406.1  GI:9524192
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 656)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished
COMMENT     Other_GSSs: RPCI-23-138J12.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 138  row: J  column: 12
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .656
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
```

```
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-138J12"
                     /sex="Female"
                     /lab_host="DH10B"
                     /clone_lib="RPCI-23"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).  "
BASE COUNT      205 a    163 c    282 g      6 t
ORIGIN

  Query Match             9.4%;  Score 94;  DB 28;  Length 656;
  Best Local Similarity   46.9%;  Pred. No. 1e-07;
  Matches  292;  Conservative    0;  Mismatches  330;  Indels    0;  Gaps    0;

Qy        223 GCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 282
              ||  ||  ||  || |||  ||  || ||  |    | ||||   |  ||  || |||
Db         32 GCAGGCAGGCAGGCAGGCAGGCAGACAGACAGGCAGGCAGACAGGCAGGCAGGCAG 91

Qy        283 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 342
              ||  || |   ||     |||| ||  || |  ||   | ||||  ||   |  || |
Db         92 GCAGGCAGGCAGGCAGACAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGACAGGCAG 151

Qy        343 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCT 402
              ||  ||  || |   |    | |||  || ||||  ||    |||   ||  || |||
Db        152 GCAGGCAGGCAGGCAGACAGGCAGGCATGTTGGCAGGCAGACAGGCAGGCAGGCAGGCAG 211

Qy        403 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 462
              |   |  ||  || |||   |   |  ||  || |||  ||  ||  ||  || | |
Db        212 ACAGACAGGCAGGCAGGCAGACAGACAGGCAGGCAGGCAGGCAGGCAGGCAGACAG 271

Qy        463 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 522
              ||   |  ||  || |||  ||   |  ||  || | |  ||  ||  ||  || | |
Db        272 GCAGACAGGCAGGCAGGCAGGCAGACAGGCAGGCAGACAGGCAGGCAGGCAGGCAGACAG 331

Qy        523 GCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGG 582
              ||   |  ||  || | |  ||  || ||  ||     |||   ||  ||  ||     |
Db        332 GCAGACAGGCAGGCAGACAGGCAGGCAGACAGGCAGACAGGCAGGCAGGCAGGCAGGCAG 391

Qy        583 GACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 642
              |  || | ||   |  | | |   | ||    || ||   ||  | ||    || |
Db        392 GCAGGCAGGCAGACAGGCAGACAGGCAGGCAGACAGGCAGGCAGGCAGACAGGCAG 451

Qy        643 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTG 702
              |   ||  | ||   || ||   ||  | ||   || ||   ||  | |||    | |
Db        452 GCAGGCAGGCAGACAGGCAGGCAGGCAGGCAGACAGGCAGGCAGGCAGGCAGGCAG 511

Qy        703 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTG 762
              |   ||  | |||   || ||   ||  | |||    | ||   ||  | |||    | |
Db        512 GCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGACAGGCAGGCAGGCAGGCAGACAG 571

Qy        763 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCT 822
              |   ||  | |||    | ||    |  | |||   || ||   ||    |||   ||
Db        572 GCAGGCAGGCAGGCAGACAGGCAGACAGGCAGGCAGGCAGGCAGGCAGACAGGCAGGCAA 631

Qy        823 GCTGGCTTGCTTGCTTGCTTGC 844
              || | |  ||  ||   |  ||
Db        632 GCAGACAGGCAGGCAGACAGGC 653


RESULT 82
BH981983/c
LOCUS       BH981983                787 bp    DNA     linear   GSS 07-OCT-2002
DEFINITION  ode73b07.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
            sequence.
ACCESSION   BH981983
VERSION     BH981983.1  GI:23513062
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 787)
  AUTHORS   Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash
            ,W., Rabinowicz,P.D. and Wilson,R.K.
  TITLE     Whole genome shotgun reads from Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: ode73  row: b  column: 07
            Seq primer: -21UPpOT forward
            Class: shotgun
            High quality sequence start: 70
            High quality sequence stop: 271.
FEATURES             Location/Qualifiers
     source          1. .787
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3712"
                     /clone_lib="B.oleracea002"
                     /note="Vector: pOTw13; Whole genome shotgun library from
                     flowering buds. DNA was purified  from a crude nuclear
                     prep using Brassica oleracea TO1000DH3 buds  provided by
                     Thosmas Osborn at the University of Wisconsin. Genomic
                     DNA was provided by Pablo Rabinowicz (CSHL) and the
                     shotgun library  prepared at Washington University Genome
                     Sequencing Center."
BASE COUNT      230 a    276 c    184 g     95 t      2 others
ORIGIN

  Query Match             9.4%;  Score 94;  DB 28;  Length 787;
  Best Local Similarity   55.5%;  Pred. No. 1.1e-07;
  Matches  181;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy        224 CCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTG 283
              | |||  | |||  |    || || |  |  |   |   || |||| | ||||||    |
Db        480 CGTGCTCGGTCGTCGCGTAGCGTGGTCGGTCGCTGGGTAGCGTGCTCGGTCGCTGCGTAG 421

Qy        284 CTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGG 343
              | ||||  |  |       || |||| | |    |   ||   ||  | | |||   | |
Db        420 CGTGCTCGGTCGCAGCGTAGCGTGCTCGGTCGCAGCGTAGCGTGCTCGGTCGCTGCGTAG 361

Qy        344 CCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTG 403
              | |||| | |    |   ||   |  | | |||   |||   |   | ||||||    |
Db        360 CGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAG 301

Qy        404 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 463
              | |||| | ||||||    || |||| | ||||||    || |||| | ||||||    |
Db        300 CGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAG 241

Qy        464 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 523
              | |||| | |||||     || |||| | ||||||    || |||| | ||||||    |
Db        240 CGTGCTCGGTCGCTCCGTAGCGTGCTCGGTCGCTGCGTAGCGTGCTCGGTCGCTGCGTAG 181

Qy        524 CTTGCTTGCTCGCTGGCCTGCTTGCT 549
              | |||| | ||||||    || ||||
Db        180 CGTGCTCGGTCGCTGCGTAGCGTGCT 155


RESULT 83
BX426372/c
LOCUS       BX426372                1115 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX426372 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
            clone CS0DG006YK02 3-PRIME, mRNA sequence.
ACCESSION   BX426372
VERSION     BX426372.1  GI:30772568
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1115)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0AG006BF01NP1.
FEATURES             Location/Qualifiers
     source          1. .1115
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DG006YK02"
                     /tissue_type="B CELLS (RAMOS CELL LINE)"
                     /cell_line="RAMOS CELL LINE"
                     /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
BASE COUNT        8 a    494 c    577 g      0 t     36 others
ORIGIN

  Query Match             9.4%;  Score 94;  DB 13;  Length 1115;
  Best Local Similarity   46.9%;  Pred. No. 1.1e-07;
  Matches  283;  Conservative    3;  Mismatches  318;  Indels    0;  Gaps    0;

Qy          1 CGCCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTG 60
              |:||||||| |  |    |  | | |||| |   |  || || |||| | ||    | :
Db        973 CSCCGCCCGCCGCGCGCGCGCGCTGCCGCGCCGGCGCCCGCCCCGCCCGCGCGTCGCGCK 914

Qy         61 GTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGG 120
                || | || |||  |   | | |||   || || |   || ||   | || ||     |
Db        913 CGCTGCCCGGCGCGCGCGCGCGCGCCCCCGCGCGCGCGCCCGCGCGCGCGCGCGCGCGCG 854

Qy        121 GGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGAC 180
               |||||  || | ||||| |       | |     |          | | ||| | |
Db        853 CGCGCGCGCGCGCGCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 794

Qy        181 AGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGC 240
                | |   ||| | ||| ||  | | |    || |  ||   ||  ||  || ||| ||
Db        793 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 734

Qy        241 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 300
                ||  ||    | |   |   ||  ||  || |||  ||  ||  ||    | |   |
Db        733 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 674

Qy        301 AGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 360
                ||  ||  ||    | |   |   ||  ||  || |||  ||  ||  ||    | |
Db        673 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 614

Qy        361 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 420
                |   |   ||  ||  ||  |   |   || ||| ||  ||  ||  :| |||  ||
Db        613 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCKCGCGCGCGC 554

Qy        421 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 480
                ||  ||  || |||  ||  ||  ||  || |||  ||  ||  ||  || |||  ||
Db        553 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 494

Qy        481 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 540
                ||  ||  || |||  ||  ||  ||  || |||  ||  ||  ||  || |||  ||
Db        493 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 434

Qy        541 CTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGA 600
                ||  ||    | |   |   ||  ||  ||  ||    | |   |  |||  ||  |
Db        433 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 374

Qy        601 TGGC 604
                ||
Db        373 GCGC 370


RESULT 84
CA056594/c
LOCUS       CA056594                734 bp    mRNA    linear   EST 04-MAR-2003
DEFINITION  ssalrgb536229 mixed_tissue Salmo salar cDNA, mRNA sequence.
ACCESSION   CA056594
VERSION     CA056594.1  GI:24386837
KEYWORDS    EST.
SOURCE      Salmo salar (Atlantic salmon)
  ORGANISM  Salmo salar
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
REFERENCE   1  (bases 1 to 734)
  AUTHORS   GRASP Consortium, Davidson,W.S., Koop,B.F. and
            http://web.uvic.ca/cbr/grasp.
  TITLE     A survey of Salmo salar transcripts from high complexity cDNA
            libraries
  JOURNAL   Unpublished
COMMENT     Contact: Koop BF
            Centre for Biomedical Research
            University of Victoria
            PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
            Tel: 250 472 4067
            Fax: 250 472 4075
            Email: bkoop@uvic.ca
            Genome Sciences Centre, BC Cancer Agency
            cDNA preparation, sequencing and bioinformatics:
            Y Butterfield, R Kirkpatrick, J Asano, N Girn, R Guin, D Lee,
            S Lee, T Olson, P Pandoh, A Prahbu, D Smailus, L Spence, J Stott,
            S Taylor, G Yang, J Schein, S Jones and M Marra.
FEATURES             Location/Qualifiers
     source          1. .734
                     /organism="Salmo salar"
                     /mol_type="mRNA"
                     /strain="McConnell"
                     /db_xref="taxon:8030"
                     /clone_lib="mixed_tissue"
                     /note="Vector: pCMVsport6; Library Creator:  Research
                     Genetics ; Atlantic salmon tissue contributors: Carlo
                     Biagi, Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.),
                     Simon Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery
                     (Crofton, B.C.), Rachel Roper (University of Victoria)"
BASE COUNT      276 a    187 c    162 g    109 t
ORIGIN

  Query Match             9.4%;  Score 93.8;  DB 14;  Length 734;
  Best Local Similarity   50.6%;  Pred. No. 1.1e-07;
  Matches  286;  Conservative    0;  Mismatches  267;  Indels   12;  Gaps    2;

Qy        299 GTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 358
              ||  | |||  |||| || |||||  | |||| ||   |   | |||  |||| ||
Db        687 GTCTGTCTGTCTGCTATGGTGACAGTCTGTCTGCCTGTCTGTCTGTCTGACTGCTATGGT 628

Qy        359 GACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 418
              ||||||  | |        ||  ||  | |||   |||    ||   |  ||   |
Db        627 GACAGTCTGTC--------TGTCTGTCTGTCTGACTGCTATGGTGACAGTCTGTCTGTCT 576

Qy        419 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 478
              | |||  ||  ||   |  |||  ||  ||   ||     ||   |  ||   ||
Db        575 GTCTGTCTGTCTGTCTGTCTGTCTGTCTGACTGCAATGGTGACAGTCTGTCTGCCT 516

Qy        479 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 538
              | |||  ||||      |   | |||  ||||      |   | |||  ||| ||   |
Db        515 GTCTGACTGCTATGGTGTCTGTCTGTCTGCTATGGTGACAGTCTGTCTGCCTGTCTGTCT 456
```

```
Qy        539 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCT 598
              | |||  ||||  || ||||||    |||  ||  ||||| || | | ||  | ||  ||
Db        455 GTCTGACTGCTATGGTGACAGT----CTGTCTGTCTGCTATGGTGTCTGTCTGTCTAACT 400

Qy        599 GATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 658
              | |    |||  |||  | ||||  | |||   ||  | |||||  | ||  |||  | ||
Db        399 GCTATGGTGACAGTCTGTCTGCCTGTCTGTCTGTCTGACTGCTATGGTGACAGTCTGTCT 340

Qy        659 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCT 718
              |   | |||   |   |       | | |   ||  | ||    | |||| |    | |
Db        339 GTCTGTCTGTCTGACTGCTATGGTGACAGTCTGTCTGTCTGTATGACTGCAATGGTGACA 280

Qy        719 GCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 778
              |   | ||||  ||  | ||  ||  | ||   ||  | |||||  | ||  |||  | ||
Db        279 GTCTGTCTGCCTGTCTGACTGCTATGGTGTCTGTCTGTCTGCTATGGTGACAGTCTGTCT 220

Qy        779 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTT 838
              ||  | |||   ||  | ||||   | ||  |||  | |||||||   |  ||||   |
Db        219 GCCTGTCTGTCTGTCTGACTGCAATGGTGACAGTCTGTCTGCCTGTCTGTCTGCTATGGT 160

Qy        839 GCTTGCTAGTGGGGCCGCTTGCCTG 863
              |  ||   ||  |      || |||
Db        159 GTCTGTCTGTCTGCTATGGTGTCTG 135


RESULT 85
CA063770/c
LOCUS       CA063770                737 bp    mRNA    linear   EST 04-MAR-2003
DEFINITION  ssalrgb530042 mixed_tissue Salmo salar cDNA, mRNA sequence.
ACCESSION   CA063770
VERSION     CA063770.1  GI:24394013
KEYWORDS    EST.
SOURCE      Salmo salar (Atlantic salmon)
  ORGANISM  Salmo salar
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
REFERENCE   1  (bases 1 to 737)
  AUTHORS   GRASP Consortium, Davidson,W.S., Koop,B.F. and
            http://web.uvic.ca/cbr/grasp.
  TITLE     A survey of Salmo salar transcripts from high complexity cDNA
            libraries
  JOURNAL   Unpublished
COMMENT     Contact: Koop BF
            Centre for Biomedical Research
            University of Victoria
            PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
            Tel: 250 472 4067
            Fax: 250 472 4075
            Email: bkoop@uvic.ca
            Genome Sciences Centre, BC Cancer Agency
            cDNA preparation, sequencing and bioinformatics:
            Y Butterfield, R Kirkpatrick, J Asano, N Girn, R Guin, D Lee,
            S Lee, T Olson, P Pandoh, A Prahbu, D Smailus, L Spence, J Stott,
            S Taylor, G Yang, J Schein, S Jones and M Marra.
FEATURES             Location/Qualifiers
     source          1. .737
                     /organism="Salmo salar"
                     /mol_type="mRNA"
                     /strain="McConnell"
                     /db_xref="taxon:8030"
                     /clone_lib="mixed_tissue"
                     /note="Vector: pCMVsport6; Library Creator:  Research
                     Genetics ; Atlantic salmon tissue contributors: Carlo
                     Biagi, Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.),
                     Simon Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery
                     (Crofton, B.C.), Rachel Roper (University of Victoria)"
BASE COUNT      274 a    185 c    163 g    115 t
ORIGIN

  Query Match             9.4%;  Score 93.8;  DB 14;  Length 737;
  Best Local Similarity   50.6%;  Pred. No. 1.1e-07;
  Matches  286;  Conservative    0;  Mismatches  267;  Indels   12;  Gaps    2;
```

```
Qy        299 GTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 358
              ||  | |||  ||||  || ||||||  | ||||  ||    |    | |||  ||||  ||
Db        706 GTCTGTCTGTCTGCTATGGTGACAGTCTGTCTGCCTGTCTGTCTGTCTGACTGCTATGGT 647

Qy        359 GACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 418
              ||||||  | |          ||   ||  | |||   |||     ||   |  ||   |
Db        646 GACAGTCTGTC--------TGTCTGTCTGTCTGACTGCTATGGTGACAGTCTGTCTGTCT 595

Qy        419 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 478
              | |||  ||  ||   |   | |||  ||  ||   ||      ||   |  ||   ||
Db        594 GTCTGTCTGTCTGTCTGTCTGTCTGTCTGACTGCAATGGTGACAGTCTGTCTGCCT 535

Qy        479 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 538
              | |||  ||||     |   | |||  ||||     |   | |||  ||| ||   |
Db        534 GTCTGACTGCTATGGTGTCTGTCTGTCTGCTATGGTGACAGTCTGTCTGCCTGTCTGTCT 475

Qy        539 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCT 598
              | |||  ||||  || ||||||    |||  ||  ||||| || | | ||  | ||  ||
Db        474 GTCTGACTGCTATGGTGACAGT----CTGTCTGTCTGCTATGGTGTCTGTCTGTCTAACT 419

Qy        599 GATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 658
              | |    |||  |||  | ||||  | |||   ||  | |||||  | ||  |||  | ||
Db        418 GCTATGGTGACAGTCTGTCTGCCTGTCTGTCTGTCTGACTGCTATGGTGACAGTCTGTCT 359

Qy        659 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCT 718
              |   | |||   |   |       | | |   ||  | ||    | |||| |    | |
Db        358 GTCTGTCTGTCTGACTGCTATGGTGACAGTCTGTCTGTCTGTATGACTGCAATGGTGACA 299

Qy        719 GCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 778
              |   | ||||  |-|  | || ||  | ||   ||  | |||||  | ||  |||  | ||
Db        298 GTCTGTCTGCCTGTCTGACTGCTATGGTGTCTGTCTGTCTGCTATGGTGACAGTCTGTCT 239

Qy        779 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTT 838
              ||  | |||   ||  | ||||   | ||  |||  | |||||||   |  ||||   |
Db        238 GCCTGTCTGTCTGTCTGACTGCAATGGTGACAGTCTGTCTGCCTGTCTGTCTGCTATGGT 179

Qy        839 GCTTGCTAGTGGGGCCGCTTGCCTG 863
              |  ||   ||  |      || |||
Db        178 GTCTGTCTGTCTGCTATGGTGTCTG 154


RESULT 86
BZ224494
LOCUS       BZ224494                1039 bp    DNA     linear   GSS 11-OCT-2002
DEFINITION  CH230-446O21.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-446O21, genomic survey sequence.
ACCESSION   BZ224494
VERSION     BZ224494.1  GI:23882852
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1039)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-446O21.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
```

```
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 446  row: O  column: 21
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .1039
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-446O21"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      265 a    229 c    278 g    267 t
ORIGIN

  Query Match             9.3%;  Score 93.2;  DB 29;  Length 1039;
  Best Local Similarity   52.2%;  Pred. No. 1.5e-07;
  Matches  303;  Conservative    0;  Mismatches  268;  Indels    9;  Gaps    4;

Qy        239 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 298
              | | |||  ||||| ||||||  ||   ||     ||  | | |||  ||||| |||
Db        456 GTCAGCTGTCTGAGATGACAGTCAGCTGTCTGAGATGCCAGTCGGCTGTCTGAGATGAC- 514

Qy        299 GTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 358
              |   | | |||  |||||  | ||||  ||   ||     ||| | | |||  |||||
Db        515 GACAGTCAGCTCTCTGAGATGCCAGTCAGCTGTCTGAGATGCTAGTCAGCTGTCTGAGAT 574

Qy        359 GACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 418
              | |||| ||    |||  ||| |||  |    ||     | || |   |  |||  |||
Db        575 GCCAGTCGGCTGTCTGAGATGCCAGTCAGCTGTCTGAGATGACTACAGTCA-GCTGTCTG 633

Qy        419 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 478
              ||   |   |||  |||   ||| || | |||  |||   ||   |   |||  |||
Db        634 AGATGACAGTCGGCTGTCTGAGATGCCTGTTGGCTGTCTGAGATGACAGTCAGCTGTCTG 693

Qy        479 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 538
              |||  |   |||| |||   |||  |   |||  |||   |||  |   |||  |||
Db        694 AGATGCCAGTCAGCTCTCTGAGATGCCAGTCAGCTGTCTGAGATGCCAGTCAGCTGTCTG 753

Qy        539 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCT 598
              ||   |  |       | | | | |||  || ||  | | ||  ||    |
Db        754 AGATGACAGTCGGCTGTCTGAGATGCCAGTTGGCTGTCTGAGATGCCTGTTGGCTGTC-- 811

Qy        599 GATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 658
              ||  |||  |||  |    |||  | ||||||   |    |||  | ||  |||  ||
Db        812 --TGAGATGACAGTCAGCTGTCTGAGATGCTAGCCAGCTGTCTGAGATGACAGTTGGCTG 869

Qy        659 GCTGGGCTGCTAGTAGGGCTGCTGGGCTG---CTAGTAGGGCTCCTGGGCTGCTAGTAGG 715
               ||| | |||  |||  |    |||  | ||      |||  |    |||  | |||  |||  ||
Db        870 TCTGAGATGCCAGTCAGCTGTCTGAGATGACTACAGTCAGCTGTCTGAGATGCCAGTCGG 929

Qy        716 GCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGG 775
              |||  | |||  |||  |    |||  | |||  |||  |    |||  | |||  |||  |
Db        930 CTGTCTGAGATGCCAGTCAGCTGTCTGAGATGCCAGTCAGCTGTCTGAGATGCCAGTCAG 989

Qy        776 GCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGG 815
              |||  | ||  |||  ||    |||  | |||  |||  ||
Db        990 CTGTCTGAGATGACAGTCGGTTGTCTGAGATGCCAGTTGG 1029


RESULT 87
DR42N9S/c
LOCUS       DR42N9S                 490 bp    DNA     linear   GSS 22-NOV-2002
DEFINITION  Danio rerio genomic clone DKEY-42N9, genomic survey sequence.
ACCESSION   AL981355
VERSION     AL981355.1  GI:25183648
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   Humphray,S.J., Huckle,E. and Hunt,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2002) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 42N9. 42N9 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene.
            Further details: http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .490
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-42N9"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
BASE COUNT      176 a    170 c     11 g    133 t
ORIGIN

  Query Match             9.3%;  Score 93;  DB 29;  Length 490;
  Best Local Similarity   50.6%;  Pred. No. 1.6e-07;
  Matches  225;  Conservative    0;  Mismatches  220;  Indels    0;  Gaps    0;

Qy        319 GACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTT 378
              |||  |   | || |||  | | |  || ||| |     ||  |   |  || | | ||
Db        489 GACTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATT 430

Qy        379 GCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 438
              | | |   |  || | | | |  || ||| ||| | | | |  || ||| ||| | | |
Db        429 GATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATT 370

Qy        439 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 498
              |  || ||| ||| | | | |  || ||| ||| | | | |  || ||| ||| | | |
Db        369 GATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATT 310

Qy        499 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 558
              |  || ||| ||| | | | |  || ||| ||| | | | |  || ||| ||| |  |||
Db        309 GATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATGGGCAGACG 250

Qy        559 GTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCT 618
              |  ||   |   |   |       ||| ||   |   |  ||||| | ||| |  ||
Db        249 GATGGATGGACGGGCAGACGGACGGATGGATGGATGGATGGATGGATGGATGGATG 190

Qy        619 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 678
              | |||   | | |  ||   | |||   | | |  ||   | |||   | | |  ||
Db        189 GATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATG 130

Qy        679 GCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 738
              | |||   | | |  ||     |||     |||     | | | ||    |   || ||
Db        129 GATGGATGGATGGATGGATGGATGGATAAATAGGTAAACAGGTTGGTGAATGAATGGACT 70

Qy        739 CCTGGGCTGCTAGTAGGGCTGCTGG 763
              |||     ||  | ||| | |||
Db         69 AATGGATGAATAAAAAGGCAGGTGG 45


RESULT 88
BX426381/c
LOCUS       BX426381                1201 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX426381 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
            clone CS0DG006YM04 5-PRIME, mRNA sequence.
ACCESSION   BX426381
VERSION     BX426381.1  GI:30784544
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

```
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 1201)
  AUTHORS  Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE    Full-length cDNA libraries and normalization
  JOURNAL  Unpublished
COMMENT    Contact: Genoscope
           Genoscope - Centre National de Sequencage
           BP 191 91006 EVRY cedex - France
           Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
           Library was constructed by Life Technologies, a division of
           Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
           http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
           Faraday Avenue Genoscope sequence ID : CS0AG006BG02QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DG006YM04"
                     /tissue_type="B CELLS (RAMOS CELL LINE)"
                     /cell_line="RAMOS CELL LINE"
                     /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
BASE COUNT        7 a    571 c    208 g    215 t    200 others
ORIGIN

  Query Match             9.3%;  Score 92.8;  DB 13;  Length 1201;
  Best Local Similarity   40.0%;  Pred. No. 1.8e-07;
  Matches  256;  Conservative   56;  Mismatches  328;  Indels    0;  Gaps    0;

Qy          5 GCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCT 64
              |: :|   :  |:|  :   :|:  |:   : : : : : :|:   : :|: :|   |
Db        660 GMGMGMGMGMGMGMGAGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGCGMGM 601

Qy         65 CCGCGCCGCCGGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCG 124
               :|:|:   :| |:| | |  :| |  :  | |:|:     |:| |  :  |:|:| |
Db        600 GMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGM 541

Qy        125 CGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGG 184
                 :   |  : :| |:|: :| | |  || |:   ||||:| |     | |  :  |
Db        540 GMGMGMGMGMGMGMGMGMGMGMGCGCGCGCGMGCGCGCGMGMGMGMGMGMGMGMGM 481

Qy        185 GACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGC 244
              |:  |:| | |:| |:  | |:|: : :| |  ||   ||  ||  || |||  ||  ||
Db        480 GMGMGMGMGMGMGMGCGMGMGMGMGCGCGCGMGCGCGCGCGCGCGCGCGCGCGCGC 421

Qy        245 TTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGC 304
                ||    | |   |   ||  ||  || ||| ||  ||  ||  ||    | |   |   ||
Db        420 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 361

Qy        305 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 364
                ||  ||    | |   |   ||  ||  || |||  ||  ||  ||    | |   |
Db        360 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 301

Qy        365 AGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 424
                |   ||  ||  ||  |   |   || |||  ||  ||  ||  || |||  ||  ||
Db        300 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 241

Qy        425 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 484
                ||  || |||  ||  ||  ||  || |||  ||  ||  ||  || |||  ||  ||
Db        240 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 181

Qy        485 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 544
                ||  || |||  ||  ||  ||  || |||  ||  ||  |:  || |||  ||  |:
Db        180 GCGCGCGCGCGCGCGCGMGCGCGCGCGCGCGCGCGCGCGCGMGMGCGCGCGCGMGCGCGM 121

Qy        545 TTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGC 604
                ||    | |:  |   |   ||  ||  |:    | |   |  |||  ||  |:  ||
```

```
Db        120 GCGCGCGMGMGMGCGCGCGCGCGCGCGCGCGMGCGCGCGCGCGCGCGCGCGCGMGCGC 61

Qy        605 TTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 644
                |: :|   |   ||  |   ||  |  |||  |  |||
Db         60 GCGMGMGCGCGCGCGCGCGCGCGCGCGCGGGGGGGGGGGG 21


RESULT 89
CB499018/c
LOCUS       CB499018                672 bp    mRNA    linear   EST 16-MAY-2003
DEFINITION  ssalkhc007023 kidney Salmo salar cDNA, mRNA sequence.
ACCESSION   CB499018
VERSION     CB499018.1  GI:29310244
KEYWORDS    EST.
SOURCE      Salmo salar (Atlantic salmon)
  ORGANISM  Salmo salar
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
REFERENCE   1  (bases 1 to 672)
  AUTHORS   GRASP Consortium, Davidson,W.S., Koop,B.F. and
            http://web.uvic.ca/cbr/grasp.
  TITLE     A survey of Salmo salar transcripts from high complexity cDNA
            libraries
  JOURNAL   Unpublished
COMMENT     Contact: Koop BF
            Centre for Biomedical Research
            University of Victoria
            PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
            Tel: 250 472 4067
            Fax: 250 472 4075
            Email: bkoop@uvic.ca
            Centre for Biomedical Research, University of Victoria cDNA
            preparation and sequencing: Roberto Alberto, Marianne Beetz-Sargent
            , Maura Busby, Peter Hunt, Linda McKinnel, BF Koop. bioinformatics:
            Gordon D Brown
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .672
                     /organism="Salmo salar"
                     /mol_type="mRNA"
                     /strain="McConnell"
                     /db_xref="taxon:8030"
                     /clone_lib="kidney"
                     /note="Vector: pBlueScriptIISK+; Library Creator:  Matthew
                     L Rise ; Atlantic salmon tissue contributors: Carlo Biagi,
                     Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.), Simon
                     Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery (Crofton,
                     B.C.), Rachel Roper (University of Victoria)"
BASE COUNT      256 a    142 c    151 g    123 t
ORIGIN

  Query Match             9.2%;  Score 92.4;  DB 14;  Length 672;
  Best Local Similarity   53.0%;  Pred. No. 2.1e-07;
  Matches  251;  Conservative    0;  Mismatches  211;  Indels   12;  Gaps    2;

Qy        303 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 362
              ||  |||| |||| ||  |    || ||| |  || ||| ||  |||| |||| ||  |
Db        663 GCTGGCTTTCTGATGGTTCCTCTGGACTGATACCTAGCTAGCTGGCTTTCTGATGGTTCC 604

Qy        363 GTAGGGCTGCTGGCTTGCTAGTAGGGCTGCT--------CGCTGGCCTGCTTGCTTGCTC 414
                 || ||| |  || ||||  ||  | ||         | |||| ||| ||  | |||
Db        603 TCTGGACTGATATCTAGCTAGCTGGCTTTCTGATGGTTCCTCTGGACTGATTCATAGCTA 544

Qy        415 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 474
              ||| || | ||        | |||| ||| || || ||| ||| || | ||        |
Db        543 GCTAGCTTTCTGATGGTTCCTCTGGACTGATTCCTAGCTAGCTAGCTTTCTGATGGTTCC 484

Qy        475 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 534
               |||| ||| || || ||| ||| || | ||        | | || ||| || || |||
Db        483 TCTGGACTGATTCCTAGCTAGCTAGCTTTCTGATGGTTCCTCCGGACTGATTCCTAGCTA 424

Qy        535 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCT 594
```

```
Db        423 GCTAGCTAGCTTTCTGATGGTTCCTCTGGACTGATTCCTAGCTAGCTAGCTTTCTGATGG 364

Qy        595 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTG----CTA 650

Db        363 TTCCTCTGGACTGATTCCTAGCTAGCTGGCTTTCTGATGGTTCCTCTGGACTGATTCCTA 304

Qy        651 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTA 710

Db        303 GCTAGCTAGCTAGCTTTCTGATGGTTCCTCTGGACTGATTCCTAGCTAGCTAGCTTTCTG 244

Qy        711 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGG 764

Db        243 ATGGTTTCTCTGGACTGATACCTAGCATGCTAGCTTTCTGATGGTTTTGCTGTG 190


RESULT 90
CNS024NL
LOCUS       CNS024NL                1152 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            235E04 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL181002
VERSION     AL181002.1  GI:7819059
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1152)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1152
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="235E04"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG235BC02SP1-end :
                     PUC-Ori"
BASE COUNT      306 a     97 c    347 g    318 t     84 others
ORIGIN

  Query Match             9.2%;  Score 92.4;  DB 29;  Length 1152;
  Best Local Similarity   53.7%;  Pred. No. 2.1e-07;
  Matches  174;  Conservative    7;  Mismatches  143;  Indels    0;  Gaps    0;

Qy        216 TACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCG 275

Db        745 TACGGGCCATGGCTGGCTGGCTGGTTGGTTGGTTGGTTGGTTGGTTTGTTGACTGG 804

Qy        276 CTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTG 335

Db        805 TTGGTTGGCTGACTGGCTGGCTGGTTGGTTGGTTGGTTGGTTGACTGGTTGGTTGGCTGG 864

Qy        336 CTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCG 395

Db        865 CTGGCTGGCTGGCTGGTTGGTTGGTTKGTTGGTTGGTTTGTTGAYTGGTTGGTTGGTTGG 924

Qy        396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455

Db        925 CTGGCTGGCTGGTTGGTTGGTTGGTTTGTTGACTGGCTGGCTGGYTGGTTGGTTTGTTGA 984

Qy        456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515

Db        985 CTGGTTGGTTGKCTGGCTGKTTKGTTTGTTGACTGGTTTGGTTGGCTKGCTGGYTGGTTG 1044

Qy        516 CTGGCCTGCTTGCTTGCTCGCTGG 539

Db       1045 GTTGGTTGGCTGGTTGTTGACTGG 1068


RESULT 91
BH731712/c
LOCUS       BH731712                  490 bp    DNA     linear   GSS 20-FEB-2002
DEFINITION  BOMAD94TF BO_2_3_KB Brassica oleracea genomic clone BOMAD94,
            genomic survey sequence.
ACCESSION   BH731712
VERSION     BH731712.1  GI:18837107
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOMAD94TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .490
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOMAD94"
                     /clone_lib="BO_2_3_KB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      100 a    167 c    174 g     49 t
ORIGIN

  Query Match             9.2%;  Score 92.2;  DB 28;  Length 490;
  Best Local Similarity   49.3%;  Pred. No. 2.2e-07;
  Matches  241;  Conservative    0;  Mismatches  248;  Indels    0;  Gaps    0;

Qy        395 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 454
```

```
Db        489 GCGTGCCCGTTCGCTACGTAGCGTGCCCGGTCGTTACGTAGCGTGCCCGGTCGCTACGTA 430

Qy        455 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 514
              ||  ||| | | |||   | ||  ||| | | |||   | ||  |||| | | |||   |
Db        429 GCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTA 370

Qy        515 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 574
              ||  ||| | | |||   | ||  ||| | | |||   |  |   |   ||  |||   |
Db        369 GCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTA 310

Qy        575 GCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAG 634
              ||      |   |       || ||   | | | ||    |  |||  ||  ||||
Db        309 GCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTA 250

Qy        635 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 694
              |  |||  ||  ||||    |  |||  ||  ||||    |  |||  ||  ||||
Db        249 GCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTA 190

Qy        695 GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 754
              |  | |  ||  ||||    |  |||  ||  ||||    |  | |  ||  ||||
Db        189 GCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTA 130

Qy        755 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 814
              |  |||  ||  ||||    |  |||  ||  ||||    |  |||  ||  ||||
Db        129 GCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCTTGCCCGGTCGCTACGTA 70

Qy        815 GGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCT 874
              |  |||| | | |||   | || |||  | | |    |  || ||||      ||| |  |
Db         69 GCGTGCCCGGTCGCTACGTAGCGTGCCCGGTCGCTACGTAGCCTGCCCCGCCCTACGTAT 10

Qy        875 GCTGTGCAG 883
                ||  | |
Db          9 CGTGCCCGG 1


RESULT 92
BH325278
LOCUS       BH325278                661 bp    DNA     linear   GSS 03-DEC-2001
DEFINITION  CH230-45J13.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
            CH230-45J13, genomic survey sequence.
ACCESSION   BH325278
VERSION     BH325278.1  GI:17255992
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 661)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 EcoRI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-45J13.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
      ►     Plate: 45  row: J  column: 13
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .661
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-45J13"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 1"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT     172 a    178 c    138 g    173 t
ORIGIN

  Query Match             9.2%;  Score 92.2;  DB 28;  Length 661;
  Best Local Similarity   73.3%;  Pred. No. 2.2e-07;
  Matches  118;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy        382 AGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 441
              | ||  | |  | ||| |||||||  |   ||| ||| |  |||||||||||| ||| ||
Db        501 AATAATGGTATTAGCTAGCCTGCTGGTGGGCTTGCTAGGTTGCTTGCTTGCTTGCTTGCT 560

Qy        442 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 501
              ||||||||||||| |||  | | |||||||  || |||  | |||||  ||||||  || ||
Db        561 TGCTTGCTTGCTTGCTTTCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTTGCTTCCTTGCT 620

Qy        502 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 542
              | |||||||||| ||| || |||||  ||||||  || || |
Db        621 TCCTTGCTTGCTTGCTTGCTTCCTTGCTTCCTTGCTT 661


RESULT 93
BX245766/c
LOCUS       BX245766                550 bp    DNA     linear   GSS 13-MAR-2003
DEFINITION  Danio rerio genomic clone DKEY-231B6, genomic survey sequence.
ACCESSION   BX245766
VERSION     BX245766.1  GI:28168100
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 231B6. 231B6 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .550
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-231B6"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
BASE COUNT     248 a    230 c     23 g     49 t
ORIGIN

  Query Match             9.2%;  Score 92;  DB 29;  Length 550;
  Best Local Similarity   53.2%;  Pred. No. 2.4e-07;
  Matches  218;  Conservative    0;  Mismatches  190;  Indels    2;  Gaps    1;

Qy        309 TTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCT--GAGGGGACAGTAG 366
              | | ||  ||   |||  |  || ||| | | | |  || ||| |  | |||  | || |
Db        549 TGGTTGTTGGTTTAGTTTGATTGGTTGGTTGGTTGGTTGGTTGGTGGGATCGGTTG 490

Qy        367 GGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 426
```

```
Db        489 GTTGGTTGGGTTGGTTGGTTGGTCGATCGGTTGGTTGGTTGGTTGGTCGATCGGTT 430

Qy        427 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 486

Db        429 GGTTGGTTGGTTGGTCGATCGGTTGGTTGGTTGGTTGTTGGTTAGTTTGATTGGTTGGTT 370

Qy        487 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 546

Db        369 GGTTGGTTGGTGGATCAGTTGGTTGGTTGGTTGTTGGTTAGTTTGATTGGTTGGTTGGTT 310

Qy        547 GCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTT 606

Db        309 GGTTGGTTGGATGATCGGTTGGTTGGTTGGTTGGATGATCGGTTGGTTGGTTGGTTGGTT 250

Qy        607 GATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 666

Db        249 GGTTGGTTGGTTGGTTGGATGATCGGTTGGTTGGTTGGTTGGATGATCGGTTGGTTGGTT 190

Qy        667 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGG 716

Db        189 GGTTGGATGATCGGTTGGTTGGTTGGTTGGATGGTTGGATGGTTGATTGG 140
```

RESULT 94
CD347167/c
LOCUS CD347167 579 bp mRNA linear EST 29-MAY-2003
DEFINITION EtESTef11c07.y1 Eimeria tenella M5-6 Excised cDNA Eimeria tenella cDNA 5' similar to TR:Q20000 Q20000 SIMILARITY WITH VON WILLEBRAND FRACTOR. [1] ;, mRNA sequence.
ACCESSION CD347167
VERSION CD347167.1 GI:31138788
KEYWORDS EST.
SOURCE Eimeria tenella
ORGANISM Eimeria tenella
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae; Eimeria.
REFERENCE 1 (bases 1 to 579)
AUTHORS Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T., Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
TITLE WashU-Merck Eimeria tenella project
JOURNAL Unpublished
COMMENT Contact: David Sibley, Ph.D.
WashU-Merck Eimeria tenella project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Contact David Sibley (toxoest@borcim.wustl.edu) for further information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Gibco
High quality sequence stop: 564.
FEATURES Location/Qualifiers
source 1. .579
/organism="Eimeria tenella"
/mol_type="mRNA"
/db_xref="taxon:5802"
/dev_stage="LS18"
/lab_host="E.coli DH10B (GeneHog, Invitrogen,Inc)"
/clone_lib="Eimeria tenella M5-6 Excised cDNA"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Sporozoites were obtained from in vitro sporulated and excysted oocysts of E.tenella grown in chickens. cDNA was synthesized from poly mRNA using an oligo-dT primer containing a XhoI site. Following second strand synthesis, EcoRI adapters were ligated to the cDNA and products were size-selected on sephacryl S500. The cDNA were ligated to EcoRI/XhoI prepared lambda ZapII (Stratagene). The primary library was mass excised using ExAssist helper phage (Stratagene). The phagemids were preciptated with PEG 8000, extracted with phenol/chloroform and electroporated into DH10B cells. The library may contain a small percentage of host or bacterial contaminants. Library materials provided by: Paul Liberator, Merck Research Labs Library constructed by: Paul Liberator, Merck Research Labs."
BASE COUNT 132 a 171 c 141 g 135 t
ORIGIN

Query Match 9.2%; Score 92; DB 14; Length 579;
Best Local Similarity 58.5%; Pred. No. 2.4e-07;
Matches 179; Conservative 0; Mismatches 125; Indels 2; Gaps 1;

```
Qy        299 GTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 358

Db        568 GCATGCATGCAGGCTTGCATGCATGCATGCGTGCTTGCATGCATGCATGCATGCATCCAG 509

Qy        359 GACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 418

Db        508 GCTTGCATGCATGCATGCGTGCTTGCATGCATGCAGGC--ACATGCATGCATGCTTGCGT 451

Qy        419 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 478

Db        450 GCATGCATGCATGCGTGCTTGCATGCATGCATGCTTGCGTGCATGCATGCTTGCATGCAT 391

Qy        479 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 538

Db        390 GCATACATGGGTGCTTGCATGCATGCATGCACGGGTGCTTGCATGCATGCACGCGTTCAC 331

Qy        539 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCT 598

Db        330 GCATGCATGGATGGGTGCGTGGACAGGTCCTTGCATGCATGCGCGCATGCATGCATGCTT 271

Qy        599 GATGGC 604

Db        270 GCATGC 265
```

RESULT 95
BH423105/c
LOCUS BH423105 622 bp DNA linear GSS 12-DEC-2001
DEFINITION BOHJE88TF BOHJ Brassica oleracea genomic clone BOHJE88, genomic survey sequence.
ACCESSION BH423105
VERSION BH423105.1 GI:17608833
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 622)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished
COMMENT Other_GSSs: BOHJE88TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .622
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOHJE88"
/clone_lib="BOHJ"

```
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       144 a    174 c    185 g    119 t
ORIGIN

  Query Match             9.2%;  Score 91.8;  DB 28;  Length 622;
  Best Local Similarity   53.8%;  Pred. No. 2.6e-07;
  Matches  189;  Conservative    0;  Mismatches  162;  Indels    0;  Gaps    0;

Qy        227 GCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTT 286
              || |||||| | |||   | ||      |   |       || |||||| | |||   |
Db        610 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 551

Qy        287 GCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 346
              ||      |   |       || || | |  | ||   ||| || | | ||||
Db        550 GCGTGCTCGGTCGCCACGTAGCGTGATCGGTCGCCACGTGGCGTGATCGGTCGCCACGTA 491

Qy        347 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 406
              || || |  |  | ||   ||  |||| | | || |    |  || ||| | |||   |
Db        490 GCGTGATCGGTCGCCACGTGGCGTGCTCGGTCGCCACGTAGCGTGTTCGGTCGCCACGTA 431

Qy        407 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 466
              || |||||| | |||   | || ||||| | |||   | || |||||| | |||   |
Db        430 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 371

Qy        467 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 526
              || |||||| | |||   | || ||||| | |||   | || |||||| | |||   |
Db        370 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTA 311

Qy        527 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCT 577
              || |||||| | |||   | ||      |   |    |  || |||| | |
Db        310 GCGTGCTCGGTCGCCACGTAGCGTGCTCGGTCGCCACGTAGCGTGCTCGGT 260


RESULT 96
AL562271
LOCUS       AL562271               1200 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL562271 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DC003YH19 3-PRIME, mRNA sequence.
ACCESSION   AL562271
VERSION     AL562271.2  GI:31286286
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1200)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12910526.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6688.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DC003CD10NP1&cluster=6688.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DC003CD10NP1.
FEATURES             Location/Qualifiers
     source          1. .1200
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DC003YH19"
                     /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT       70 a    371 c    442 g    129 t    188 others
ORIGIN

  Query Match             9.2%;  Score 91.8;  DB 9;  Length 1200;
  Best Local Similarity   40.5%;  Pred. No. 2.7e-07;
  Matches  286;  Conservative   50;  Mismatches  371;  Indels    0;  Gaps    0;

Qy         86 CGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGACCGAGGGCCCCGGGAGAA 145
              || |  |  | ||||    ||| |  |  ||||| |       |    |  | || | |
Db          2 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 61

Qy        146 CGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTG 205
              || | |  || |    |||| | |     | |  |  | |   ||| | ||| ||  | |
Db         62 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 121

Qy        206 AGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCC 265
              |    || |  ||   ||  ||  || ||| ||  ||  ||    | |   |  ||
Db        122 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 181

Qy        266 TGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTA 325
              ||  || ||| ||  ||  ||    | |   |   ||  ||  ||    | |   |
Db        182 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 241

Qy        326 GGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTA 385
              ||  ||  || |||  |  |||  ||    | |   |   |   ||  ||  ||  |
Db        242 CGCGCGCGCGCGCGCGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 301

Qy        386 GGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 445
              |   || |||  ||  ||  || :|| |||  || :||  ||  |: |||  ||  ||
Db        302 CGCGCGCGCGCGCGCGCGCGCGCGYGCGCGCGCGCGYGCGCGCGCGYGCGCGCGCGCG 361

Qy        446 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 505
              |: :|: ||| · ||  ||  ||  || |||  |:  ||  ||  :|: |||  |:  ||
Db        362 CGYGYGYGCGCGCGCGCGCGCGCGCGCGCGCGCGYGCGCGCGCGYGMGCGCGYGYGCGCG 421

Qy        506 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGC 565
              || ||: :||  ||  ||  ||  |: :|:  |:  |:  ||    | |:  ||  |
Db        422 CGCGTGYGYGCGCGCGCGCGCGCGYGYGYGCGYGMGYGCGCGCGCGYGHGTGTGYGHR 481

Qy        566 TGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGC 625
              : | : : ||:  : | |  ||  | :| :|:|   |  ||  ::|  :  ||:  |
Db        482 YACRYAYGTGYGYRTGTGYAYGTKTGTMTMHCYGTGTGTGTGYGMSTGWSTGTGYGTGTG 541

Qy        626 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGC 685
              ||: :|:    |  |   ||: :  |  | :||   |  ||   |  |   :||  |
Db        542 TGHGMGYGTTTTTTTTTTHCTKTKTTTKTGTSTGTGTGTGTGTGTGCGCGCGYGCGCGCG 601

Qy        686 TGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGC 745
              |   |  ||: |::  |   :|  |:  |   |   |  ||:  ||  |  ::   |
Db        602 CGTGYGCGGGKGTVYGTGTGCSCGYGYGCGCGCGTGTGTGTGYGYGTGTGTGYSTGTGTG 661

Qy        746 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGT 792
              ||   ||  |  ||   |  :||  |:  :  :|:  |  ||   ||
Db        662 TGTGTGTGTGYGTGTGTGTGYGCGYGYGYRTGYGYRTGTGTGTGTGT 708


RESULT 97
CB811482
LOCUS       CB811482               404 bp    mRNA    linear   EST 16-MAY-2003
DEFINITION  AMGNNUC:MRPE4-00166-A1-A mrpe4 (10380) Rattus norvegicus cDNA clone
            mrpe4-00166-a1 5', mRNA sequence.
ACCESSION   CB811482
VERSION     CB811482.1  GI:29934380
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 404)
  AUTHORS   Amgen EST Program.
```

```
  TITLE     Amgen Rat EST Program
  JOURNAL   Unpublished
COMMENT     Contact: Dan Fitzpatrick
            Amgen, Inc
            One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
            Tel: 805 447-4881
            Plate: 00166  row: a  column: 1.
FEATURES             Location/Qualifiers
     source          1. .404
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="mrpe4-00166-a1"
                     /tissue_type="placenta embryo"
                     /clone_lib="mrpe4 (10380)"
                     /note="Vector: pSPORT1; Site_1: Sal1; Site_2: NotI;
                     placenta embryo day 17"
BASE COUNT       36 a    154 c    102 g    110 t      2 others
ORIGIN

  Query Match             9.1%;  Score 91.4;  DB 14;  Length 404;
  Best Local Similarity   60.1%;  Pred. No. 3.1e-07;
  Matches  152;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy        387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 446
              | ||||  ||  ||||||| ||| ||| ||  |||||| ||| |||  ||  |||||| |
Db        152 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 211

Qy        447 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 506
              || |||  ||  |||||| ||  |||  |   |||||  ||| |||  ||   ||||| |
Db        212 GCCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTCCCTGCCT 271

Qy        507 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 566
              |  |||  |   |||| | ||| ||   ||  |||||| |||      | | ||  | ||
Db        272 GTCTGCCTGTCTGCCTCCCTGCCTGTCTGCCTGCCTGCCTGCCTCCCTGCCTGTCTGCCT 331

Qy        567 GCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCT 626
              || ||| |||   |  | | |    ||||||||  || ||   |   | |||   | ||
Db        332 GCCTGCCTGCCTGGCTGCCTGCCTCCCTGCCTGTCTGCCTGCCTGCCTGCCTGGCTGCCT 391

Qy        627 GCTAGTAGGGCTG 639
              ||  |   | |||
Db        392 GCCTGGCTGCCTG 404


RESULT 98
BI688458/c
LOCUS       BI688458                745 bp    mRNA    linear   EST 18-SEP-2001
DEFINITION  603316395F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5356246 5',
            mRNA sequence.
ACCESSION   BI688458
VERSION     BI688458.1  GI:15651087
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 745)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11905  row: n  column: 23
            High quality sequence stop: 650.
FEATURES             Location/Qualifiers
     source          1. .745
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5356246"
                     /sex="female, virgin"
                     /tissue_type="infiltrating ductal carcinoma"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam6"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      280 a    140 c    187 g    138 t
ORIGIN

  Query Match             9.1%;  Score 91.4;  DB 12;  Length 745;
  Best Local Similarity   77.2%;  Pred. No. 3.1e-07;
  Matches  149;  Conservative    0;  Mismatches   41;  Indels    3;  Gaps    3;

Qy        385 AGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 444
              |||  |||| ||| || ||||||||||||| ||| ||||| ||| ||||| ||| || |||
Db        320 AGGTTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTG-TTGTTTGCTTGCTTGCTTGC 262

Qy        445 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 504
              |||| |||| ||| || |||||||||||| ||| | |||| || | | | ||| || |||
Db        261 TTGC-TGCTTGCTTGCTTGCTTGCTTGCTTG-CTGCCTGTTCGTTTGCTTGCTTGC 204

Qy        505 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGG 564
              ||||||||| ||| || |||||||||||| ||| || ||||||||     |   ||  |
Db        203 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGTTCGTTTGC 144

Qy        565 CTGCTTGCTTGCT 577
              |||||||||||||
Db        143 TTGCTTGCTTGCT 131


RESULT 99
FR0038988/c
LOCUS       FR0038988               461 bp    DNA     linear   GSS 22-OCT-1999
DEFINITION  Fugu rubripes GSS sequence, clone 089D07bA10, genomic survey
            sequence.
ACCESSION   AL126486
VERSION     AL126486.1  GI:6108101
KEYWORDS    GSS; genome survey sequence.
SOURCE      Takifugu rubripes
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K.,
            Umrania,Y., Williams,G. and Brenner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
            Centre, Hinxton, Cambridge, CB10 1SB. UK Email:
            biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Takifugu rubripes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:31033"
                     /clone="089D07bA10"
                     /clone_lib="cosmid 089D07"
```

```
BASE COUNT      112 a    170 c     80 g     70 t     29 others
ORIGIN

  Query Match             9.1%;  Score 91.2;  DB 29;  Length 461;
  Best Local Similarity   53.9%;  Pred. No. 3.4e-07;
  Matches  165;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy        364 TAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 423
              ||||    |||||| | |    ||   | | | |     ||| | |   |  |||||  |
Db        426 TAGGTACCCTGGCTGGNTGAGTGGCTGGNTAGGTACNTCGCTAGGTGAGTGNCTGGCTAG 367

Qy        424 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 483
               | ||| ||| ||||    | |    |  || ||| |   ||| ||| | | |||| |  |
Db        366 GTAGCTGGCTGGCTGNNTGGGTGAGTGACTGGCTAGGNAGCTGGCTAGGTGGCTGNCTAG 307

Qy        484 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 543
               |   | | | |||| |  | |    | ||| |||||     | | | | | ||||||  |
Db        306 GTGNNTAGGTAGCTGNCTGGGTGAGTGGCTGGCTGGGTGAGTGGNTAGNTAGCTGGCTAG 247

Qy        544 CTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGG 603
               | ||||    |  ||  ||   | | ||| |||| | ||   |   ||  ||  | |||
Db        246 NTAGCTGGCTAGNTAGGTGGCTAGGTGGCTGGCTAGGTGGCTAGGTTGCTGGCNAGGTGG 187

Qy        604 CTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 663
              || | |||  ||   | | ||  |||||  ||   ||| ||  |||||   ||  ||| |
Db        186 CTAGGTAGGTGGCTAGGTAGGTGGCTAGGTGGCTGGCTAGGTGGCTAGGTAGGTGGCTAG 127

Qy        664 GCTGCT 669
              |   ||
Db        126 GTNNCT 121


RESULT 100
BH729559
LOCUS       BH729559                485 bp    DNA     linear   GSS 20-FEB-2002
DEFINITION  BOMAZ10TR BO_2_3_KB Brassica oleracea genomic clone BOMAZ10,
            genomic survey sequence.
ACCESSION   BH729559
VERSION     BH729559.1  GI:18834954
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 485)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .485
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOMAZ10"
                     /clone_lib="BO_2_3_KB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT     80 a    142 c    155 g    108 t
ORIGIN

  Query Match             9.1%;  Score 90.8;  DB 28;  Length 485;
  Best Local Similarity   54.7%;  Pred. No. 4e-07;
  Matches  210;  Conservative    0;  Mismatches  162;  Indels   12;  Gaps    1;

Qy        407 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 466
              ||| | |||||     || |||  | |||||     || |||| | |||||     || |
Db        108 GCTCGGTCGCTACGTAGCGTGCACGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGT 167

Qy        467 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 526
              ||| | |||||     || |||| | |||||     || |||| | |||||     ||||
Db        168 GCTCGGTCGCTACGAAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCTT 227

Qy        527 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACG 586
              ||| | |||||     |||||||  |  |  |     |  |||| | | |||| |
Db        228 GCTCGGTCGCTACGTAGCTTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACG------ 281

Qy        587 GTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 646
                    || |||| | | | ||    |  |||| ||  ||||    |  |||| ||
Db        282 ------TAGCGTGATCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTC 335

Qy        647 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCT 706
              ||||    |  |||| ||  ||||    |  |||| ||  ||||    |  | || ||
Db        336 GCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTC 395

Qy        707 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCT 766
              ||||    |  |||| ||  ||||    |  | || ||  ||||    |  |||| ||
Db        396 GCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTCGCTACGTAGCGTGCTCGGTC 455

Qy        767 GCTAGTAGGGCTGCTGGGCTGCTA 790
              ||||       |||| ||  ||||
Db        456 GCTACGTAACGTGCTCGGTCGCTA 479


Search completed: August 22, 2003, 00:39:05
Job time : 2086.54 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 21, 2003, 18:39:57 ; Search time 282.141 Seconds
(without alignments)
9577.262 Million cell updates/sec

Title: US-09-831-000-1_COPY_132719_133719
Perfect score: 1001
Sequence: 1 cgccgcccggcatggcggtc.....cgggcacccggcccagatct 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1001	100.0	128139	24	AAI64291	RRV genome nucleot
2	1001	100.0	133719	21	AAC64754	Macaca mulatta rha
3	193.4	19.3	576	22	ABA41262	Probe #19728 for g
4	193.4	19.3	576	22	AAI57480	Probe #26166 used
5	174.6	17.4	555	22	ABA31139	Probe #9605 for ge
6	174.6	17.4	555	22	AAI44102	Probe #12788 used
c 7	134.2	13.4	133719	21	AAC64754	Macaca mulatta rha
c 8	127.6	12.7	35100	20	AAV73802	KSHV LUR DNA (nucl
c 9	127.6	12.7	137507	19	AAV19941	KSHV long unique c
c 10	124.8	12.5	128139	24	AAI64291	RRV genome nucleot
c 11	98.4	9.8	385	16	AAQ95095	Configuration dete
c 12	96.8	9.7	344	16	AAQ84040	Plasmid pCU19. Sy
13	96.8	9.7	385	16	AAQ95095	Configuration dete
14	96.4	9.6	14212	22	AAK79692	Human immune/haema
15	96.4	9.6	14212	22	AAK80024	Human immune/haema
16	96.4	9.6	14212	22	AAK85921	Human immune/haema
17	95.2	9.5	344	16	AAQ84040	Plasmid pCU19. Sy
18	93.8	9.4	1172	24	ABQ31996	Oligonucleotide fo
c 19	93.8	9.4	1172	24	ABQ31997	Oligonucleotide fo
c 20	92.2	9.2	1172	24	ABQ31994	Oligonucleotide fo
21	92.2	9.2	1172	24	ABQ31995	Oligonucleotide fo
c 22	91.6	9.2	2130	25	ABZ39851	N. gonorrhoeae nuc
c 23	90.8	9.1	1172	24	ABQ31996	Oligonucleotide fo
24	90.8	9.1	1172	24	ABQ31997	Oligonucleotide fo
25	90.6	9.1	1172	24	ABQ31994	Oligonucleotide fo
c 26	90.6	9.1	1172	24	ABQ31995	Oligonucleotide fo
c 27	87.2	8.7	2855	22	ABA15389	Human nervous syst
28	79.8	8.0	1830121	17	AAT42063	Haemophilus influe
c 29	78.8	7.9	15682	22	AAK89506	Human digestive sy
c 30	78	7.8	169	24	AAL40273	BASB204 related DN
31	77.6	7.8	7202	22	ABA19579	Human nervous syst
32	77.6	7.8	7202	22	AAK71750	Human immune/haema
33	77	7.7	42999	24	ABS65032	Invertebrate forag
34	75.4	7.5	42999	25	ABX11086	Human ribosomal RN
c 35	74.8	7.5	428	24	ABQ97101	Mouse ES cell rela
c 36	74.8	7.5	436	24	ABQ97086	Mouse ES cell rela
37	74.2	7.4	24512	22	ABA19578	Human nervous syst
38	74.2	7.4	24512	22	AAK67571	Human immune/haema
39	74.2	7.4	24512	22	AAK71749	Human immune/haema
c 40	73.2	7.3	388	24	ABQ97027	Mouse ES cell rela
c 41	73.2	7.3	15720	22	AAI29504	C899P determined c
c 42	73.2	7.3	15720	24	ABL62358	Colon adenocarcino
c 43	73.2	7.3	15720	25	ABZ33690	Human colon tumour
c 44	72.2	7.2	409	24	ABQ97094	Mouse ES cell rela
45	71.8	7.2	16084	22	AAK69442	Human immune/haema
c 46	71.8	7.2	16084	22	AAK82881	Human immune/haema
c 47	71.4	7.1	392	24	ABQ97468	Mouse ES cell rela
48	71.4	7.1	6286	22	AAS46592	Tumour suppressor
c 49	71.2	7.1	3588	23	AAS70903	DNA encoding novel
c 50	71.2	7.1	3588	23	AAS92599	DNA encoding novel
51	71.2	7.1	7614	22	AAK85846	Human immune/haema
c 52	70.8	7.1	157	24	AAL40272	BASB204 related DN
c 53	70.8	7.1	379	24	ABQ96978	Mouse ES cell rela
c 54	70.8	7.1	427	24	ABQ96982	Mouse ES cell rela
c 55	70.6	7.1	410	24	ABQ97057	Mouse ES cell rela
56	69.4	6.9	368	22	AAL06216	Human reproductive
57	69.4	6.9	368	22	AAL07359	Human reproductive
58	68.8	6.9	600	6	AAN50853	Sequence encoding
59	68.8	6.9	13321	22	AAS46422	Tumour suppressor
60	67.2	6.7	114955	20	AAX53491	Human adenosine A1
61	67	6.7	397	24	ABQ97197	Mouse ES cell rela
c 62	66.8	6.7	183	24	AAL40271	BASB204 related DN
c 63	66.8	6.7	186	24	AAL40264	BASB204 related DN
c 64	66.8	6.7	228	24	AAL40265	BASB204 related DN
c 65	66.8	6.7	240	24	AAL40268	BASB204 related DN
c 66	66.8	6.7	252	24	AAL40266	BASB204 related DN
c 67	66.8	6.7	264	24	AAL40267	BASB204 related DN
c 68	66.8	6.7	267	24	AAL40269	BASB204 related DN
c 69	66.8	6.7	267	24	AAL40270	BASB204 related DN
c 70	66.6	6.7	32249	22	ABA17155	Human nervous syst
c 71	66	6.6	147	25	ABZ40780	N. gonorrhoeae nuc
72	65.4	6.5	3249	23	AAS76504	DNA encoding novel
c 73	65.2	6.5	368	20	AAV88624	EST clone GX576.
c 74	64.2	6.4	1771	18	AAT62138	Leishmania tropica
c 75	64.2	6.4	1771	19	AAV47558	Leishmania antigen
c 76	64.2	6.4	1771	20	AAZ25621	Leishmania tropica
c 77	64.2	6.4	1771	24	AAF88524	L. tropica LT-210
c 78	64.2	6.4	1771	24	AAD40285	Leishmania tropica
c 79	64.2	6.4	1771	24	ABK81733	Leishmania antigen
c 80	64.2	6.4	1771	24	AAS96022	Leishmania antigen
81	63.4	6.3	8174	24	AAL44951	Equine herpesvirus

c	82	63	6.3	1337	20	AAZ17263	Human gene express
c	83	62.8	6.3	149	24	AAL40260	BASB204 related DN
c	84	62.8	6.3	149	24	AAL40261	BASB204 related DN
c	85	62.8	6.3	149	24	AAL40262	BASB204 related DN
c	86	62.8	6.3	149	24	AAL40263	BASB204 related DN
c	87	62.8	6.3	2084	24	AAL40256	BASB204 related DN
c	88	62.8	6.3	2084	24	AAL40257	BASB204 related DN
	89	62.8	6.3	31718	22	AAS39915	Genomic sequence #
	90	62.8	6.3	31718	22	AAS39916	Genomic sequence #
	91	62.8	6.3	31718	22	AAL06415	Human reproductive
	92	62.8	6.3	31718	22	AAL06416	Human reproductive
c	93	62.8	6.3	31718	22	AAK73104	Human immune/haema
c	94	62.8	6.3	31718	22	AAK73120	Human immune/haema
	95	62.8	6.3	31718	22	AAK87442	Human immune/haema
	96	62.8	6.3	31718	22	AAK87443	Human immune/haema
c	97	62.8	6.3	31718	22	AAK87573	Human immune/haema
c	98	62.8	6.3	31718	22	AAK87592	Human immune/haema
	99	62.8	6.3	31718	22	AAK90359	Human digestive sy
	100	62.8	6.3	31718	22	AAK90360	Human digestive sy
	101	61.2	6.1	15682	22	AAK89506	Human digestive sy
	102	61	6.1	223	22	AAL36820	Human musculoskele
	103	61	6.1	223	22	AAK64806	Human immune/haema
	104	61	6.1	223	25	ABX59808	cDNA encoding nove
c	105	61	6.1	1542	10	AAN90025	DNA encoding antig
	106	61	6.1	1951	24	ABQ19254	Oligonucleotide fo
c	107	61	6.1	1951	24	ABQ19255	Oligonucleotide fo
c	108	60.8	6.1	1951	24	ABQ19254	Oligonucleotide fo
	109	60.8	6.1	1951	24	ABQ19255	Oligonucleotide fo
	110	60.8	6.1	2239	24	ABK90515	Human pyroglutamyl
	111	60.6	6.1	230	22	AAK54984	Human immune/haema
	112	60.6	6.1	1951	24	ABQ19256	Oligonucleotide fo
c	113	60.6	6.1	1951	24	ABQ19257	Oligonucleotide fo
	114	60.6	6.1	6109	24	ABL32327	Human immune syste
	115	60.6	6.1	6109	24	AAS61078	Human gene regulat
	116	60.4	6.0	154746	24	AAD25519	Human herpesvirus
c	117	60.4	6.0	154746	24	AAD25519	Human herpesvirus
c	118	60.4	6.0	168174	24	ABT11173	Human 5-lipoxygena
c	119	60.4	6.0	168273	24	ABT11114	Human 5-lipoxygena
c	120	60.2	6.0	1412	24	AAL46206	Human liver cancer
	121	59.6	6.0	8129	22	AAS46763	Tumour suppressor
	122	59.4	5.9	223	22	AAL36821	Human musculoskele
	123	59.4	5.9	223	22	AAK64807	Human immune/haema
	124	59.4	5.9	223	25	ABX59809	cDNA encoding nove
c	125	58.8	5.9	10483	22	AAK80686	Human immune/haema
c	126	58.6	5.9	393	24	ABQ97710	Mouse ES cell rela
c	127	58.4	5.8	148834	24	ABK83570	Human cDNA differe
	128	58.2	5.8	12733	24	ABK98631	Vector pEPEF14 con
	129	58.2	5.8	12739	24	ABK98592	Vector pEPEF1 cont
c	130	57.2	5.7	6171	24	ABL32789	Human immune syste
c	131	57.2	5.7	32187	22	AAK89946	Human digestive sy
	132	57	5.7	194	20	AAX01333	Allelic ladder, HU
	133	57	5.7	600	24	ABQ52496	Oligonucleotide fo
c	134	57	5.7	600	24	ABQ52497	Oligonucleotide fo
c	135	56.8	5.7	3595	20	AAX08988	Sequence encoding
c	136	56.8	5.7	3595	25	ABX12095	Human cerberus-lik
	137	56.8	5.7	18376	22	AAK67167	Human immune/haema
c	138	56.6	5.7	412	22	AAS38476	Novel human diagno
c	139	56	5.6	573	22	ABA64355	Human foetal liver
c	140	56	5.6	573	22	ABA76635	Human foetal liver
c	141	56	5.6	573	22	ABA31495	Probe #9961 for ge
c	142	56	5.6	573	22	ABA41143	Probe #19609 for g
c	143	56	5.6	573	22	AAK12816	Human brain expres
c	144	56	5.6	573	22	AAK25269	Human brain expres
c	145	56	5.6	573	22	AAK38543	Human bone marrow
c	146	56	5.6	573	22	AAK51276	Human bone marrow
c	147	56	5.6	573	24	ABS12615	Human genome-deriv
c	148	56	5.6	573	24	ABS24815	Human genome-deriv
c	149	56	5.6	1951	24	ABQ19256	Oligonucleotide fo
	150	56	5.6	1951	24	ABQ19257	Oligonucleotide fo
c	151	56	5.6	45546	20	AAX23520	Human kidney amino
c	152	55.8	5.6	159095	25	ABZ80818	Human PAI-2 polymo
c	153	55.6	5.6	137	24	AAL40258	BASB204 related DN
c	154	55.6	5.6	137	24	AAL40259	BASB204 related DN
c	155	55.6	5.6	354	24	ABQ97167	Mouse ES cell rela
c	156	55.6	5.6	1960	18	AAT96724	Brassica sp. tapet
c	157	55.6	5.6	3496	20	AAZ27622	Plasmid TOPI-1. S
c	158	55.6	5.6	3766	20	AAZ27621	Plasmid TOP-1. Sy
c	159	55.6	5.6	4936	20	AAZ27618	Plasmid TOG-1. Sy
c	160	55	5.5	186	21	AAC14822	Human secreted pro
	161	54.6	5.5	1218	21	AAA02488	Human colon cancer
	162	54.4	5.4	160820	25	ABQ76673	Androgen receptor
c	163	53.8	5.4	345	24	ABQ96922	Mouse ES cell rela
	164	53.8	5.4	356	23	AAS76990	DNA encoding novel
c	165	53.8	5.4	789	20	AAX40044	Prostate cancer as
c	166	53.8	5.4	2409	22	AAH17441	Human cDNA sequenc
c	167	53.8	5.4	4065	22	AAD16913	Human phorbol acti
	168	53.6	5.4	849	24	ABQ29960	Oligonucleotide fo
c	169	53.6	5.4	849	24	ABQ29961	Oligonucleotide fo
c	170	53.2	5.3	26166	25	ACC50869	Human secreted pro
c	171	53.2	5.3	26166	25	ABZ71491	Secreted protein g
	172	53.2	5.3	34634	24	AAD31198	Human WKL-1 genomi
c	173	53.2	5.3	97835	24	ABK84796	Human cDNA differe
	174	52.8	5.3	712	24	ABQ40858	Oligonucleotide fo
c	175	52.8	5.3	712	24	ABQ40859	Oligonucleotide fo
	176	52.8	5.3	754	24	ABQ40800	Oligonucleotide fo
c	177	52.8	5.3	754	24	ABQ40801	Oligonucleotide fo
c	178	52.6	5.3	319608	21	AAH51601	Human chromosome 1
c	179	52.6	5.3	319608	22	AAS09301	Human schizophreni
	180	52.4	5.2	190	20	AAX01332	Allelic ladder, HU
c	181	52.4	5.2	350	16	AAQ95194	Simple tandem repe
	182	52.2	5.2	178	20	AAX01330	Allelic ladder, HU
	183	52.2	5.2	678	24	ABQ28082	Oligonucleotide fo
c	184	52.2	5.2	678	24	ABQ28083	Oligonucleotide fo
	185	52.2	5.2	1000	21	AAA02484	Human colon cancer
c	186	52.2	5.2	2408	21	AAA28523	Human opioid growt
c	187	52.2	5.2	3250	25	ABZ79881	Human nucleic-acid
	188	52	5.2	7597	24	ABL33012	Human immune syste
	189	51.8	5.2	2218	14	AAQ47929	Paired basic amino
	190	51.8	5.2	4361	24	AAS94758	Human DNA sequence
	191	51.8	5.2	4403	14	AAQ47927	Paired basic amino
	192	51.8	5.2	4403	24	ABN95905	Gene #2403 used to
c	193	51.6	5.2	538	16	AAQ95198	Simple tandem repe
	194	51.4	5.1	182	20	AAX01331	Allelic ladder, HU
c	195	51.4	5.1	72604	20	AAZ10752	Genomic sequence o
c	196	51.4	5.1	72604	24	ABK43231	Human HKNG1 gene.
	197	51.4	5.1	98472	24	ABS54549	Human multidrug re
c	198	51	5.1	369	24	ABL82290	Human ovarian canc
c	199	51	5.1	497	21	AAA82163	N. meningitidis pa
c	200	51	5.1	3092	23	AAS69141	DNA encoding novel
c	201	50.8	5.1	12007	24	ABL32716	Human immune syste
	202	50.6	5.1	22655	22	AAK70122	Human immune/haema
	203	50.4	5.0	405	20	AAX89576	Autonomously repli
	204	50.4	5.0	405	21	AAL44929	Multiple tandem in
c	205	50.4	5.0	8459	22	AAC89557	Human histone deac
c	206	50.4	5.0	8460	24	ABK87718	Human cDNA encodin
c	207	50.2	5.0	2000	24	ABZ17227	Arabidopsis thalia
c	208	50.2	5.0	155074	24	ABN85735	Human genomic regi
c	209	50	5.0	200	23	ABV61578	Human prostate exp
c	210	50	5.0	1359	20	AAZ17254	Human gene express
	211	50	5.0	7803	20	AAX55272	Human receptor-rel
	212	50	5.0	7814	22	AAS46529	Tumour suppressor
c	213	50	5.0	10577	21	AAA81494	N. meningitidis pa
	214	50	5.0	22400	25	ABT16305	Zinc finger protei
c	215	50	5.0	349980	21	AAF21610	Neisseria meningit
c	216	50	5.0	1437668	21	AAA81490	N. meningitidis B
c	217	49.8	5.0	348	18	AAT65077	Canine genomic mic
c	218	49.8	5.0	348	20	AAX17814	Microsatellite rep
	219	49.8	5.0	6292	22	AAS46735	Tumour suppressor
c	220	49.8	5.0	9831	22	ABA07250	Human pancreatic c
c	221	49.8	5.0	9831	22	AAK89891	Human digestive sy
c	222	49.8	5.0	113033	25	AAL54213	SR protein-specifi
	223	49.6	5.0	4182	22	AAH26524	Human proton/oligo
	224	49.6	5.0	26555	22	AAK68372	Human immune/haema
	225	49.6	5.0	26555	22	AAK68605	Human immune/haema
	226	49.6	5.0	26555	22	AAI62833	Human genomic DNA
	227	49.6	5.0	136328	24	ABZ35015	Human gene express

c 228	49.4	4.9	519	22	ABA61017	Human foetal liver
c 229	49.4	4.9	519	22	AAK09310	Human brain expres
c 230	49.4	4.9	519	22	AAK35199	Human bone marrow
c 231	49.4	4.9	519	22	AAI40915	Probe #9601 used t
c 232	49.4	4.9	519	23	ABS34947	Human liver single
233	49.4	4.9	3593	22	AAD02774	Murine B7-4 protei
234	49.4	4.9	3593	22	AAD02709	Murine B7-4 protei
235	49.4	4.9	3593	25	AAD51099	Murine B7-4 DNA.
236	49.4	4.9	3593	25	ABT14004	Murine PD-L1 codin
237	49.4	4.9	12237	24	ABL34359	Human immune syste
238	49.2	4.9	249487	24	ABN85733	Mouse genomic regi
239	49	4.9	676	24	ABQ42364	Oligonucleotide fo
c 240	49	4.9	676	24	ABQ42365	Oligonucleotide fo
241	49	4.9	1189	25	ABT33357	NOVX DNA sequence
242	49	4.9	1416	25	ABZ20967	Animal test kit ma
c 243	49	4.9	2348	21	AAA28525	Human opioid growt
c 244	48.8	4.9	650	16	AAT01589	Rat neu promoter.
c 245	48.8	4.9	1682	22	AAD20196	Human translation
c 246	48.8	4.9	1682	22	AAF86072	Human translation
c 247	48.8	4.9	1682	24	ABK83618	Human cDNA differe
c 248	48.8	4.9	2561	22	AAH26500	Rabbit low density
249	48.8	4.9	7311	22	AAD07025	Pseudomonas aerugi
c 250	48.8	4.9	27048	22	AAS02354	Porcine Gal alpha(
251	48.6	4.9	174	20	AAX01329	Allelic ladder, HU
c 252	48.6	4.9	10462	24	ABL56820	Human homeobox pro
253	48.4	4.8	120	16	AAQ84038	Plasmid pCU19 olig
c 254	48.4	4.8	120	16	AAQ84038	Plasmid pCU19 olig
255	48.4	4.8	120	16	AAQ95093	Configuration dete
c 256	48.4	4.8	120	16	AAQ95093	Configuration dete
257	48.4	4.8	462586	25	ABQ84281	Chromosome 2 Asthm
258	48.2	4.8	170	20	AAX01328	Allelic ladder, HU
c 259	48.2	4.8	381	23	AAS74660	DNA encoding novel
c 260	48.2	4.8	1623	25	ABZ75087	Mouse embryonic st
261	48.2	4.8	7045	24	ABS97907	Human UDP-glucuron
262	48.2	4.8	57728	22	AAC87588	Human 9p11 chromos
c 263	48.2	4.8	397658	25	ABQ83210	Human transporter
264	48	4.8	1127	21	AAA02477	Human colon cancer
265	48	4.8	2714	22	AAS41712	Genomic sequence #
266	48	4.8	157875	24	ABK99972	Human CADPKL genom
c 267	47.8	4.8	418	24	ABK89778	Plasmid for signal
268	47.8	4.8	687	24	ABQ19226	Oligonucleotide fo
c 269	47.8	4.8	687	24	ABQ19227	Oligonucleotide fo
270	47.8	4.8	3228	22	AAC90434	Murine Lymph node
271	47.8	4.8	201143	24	ABK83568	Human DNA differen
c 272	47.6	4.8	572	24	ABQ47268	Oligonucleotide fo
273	47.6	4.8	572	24	ABQ47269	Oligonucleotide fo
c 274	47.6	4.8	836	25	ABZ75079	Mouse embryonic st
275	47.6	4.8	2779	25	ABZ34851	Coding sequence SE
c 276	47.6	4.8	8482	23	ABL13462	Drosophila melanog
277	47.6	4.8	13304	25	AAD51620	Human cytochrome P
278	47.6	4.8	14776	24	ABK83747	Human cDNA differe
279	47.6	4.8	14776	24	ABN95590	Gene #2088 used to
280	47.6	4.8	659158	25	ABX16390	Mouse high growth
281	47.4	4.7	585	24	ABQ50506	Oligonucleotide fo
c 282	47.4	4.7	585	24	ABQ50507	Oligonucleotide fo
c 283	47.4	4.7	1312	20	AAZ17264	Human gene express
c 284	47.4	4.7	2289	21	AAA28526	Human opioid growt
c 285	47.4	4.7	2290	21	AAA28522	Human opioid growt
c 286	47.4	4.7	4139	24	ABN95623	Gene #2121 used to
c 287	47.4	4.7	4139	24	ABL67071	Thyroid cancer rel
c 288	47.4	4.7	4139	24	ABL67544	Thyroid cancer rel
c 289	47.4	4.7	4139	24	ABL68884	Kidney cancer rela
c 290	47.4	4.7	4139	24	ABK09797	Human ovarian tumo
c 291	47.4	4.7	4139	25	ACC50234	Breast cancer asso
292	47.4	4.7	6419	24	ABL32266	Human immune syste
293	47.4	4.7	72604	20	AAZ10752	Genomic sequence o
294	47.4	4.7	72604	24	ABK43231	Human HKNG1 gene.
295	47.2	4.7	320	21	AAA38185	Primer used in the
296	47.2	4.7	668	24	ABQ40862	Oligonucleotide fo
c 297	47.2	4.7	668	24	ABQ40863	Oligonucleotide fo
298	47.2	4.7	1756	25	ABZ73862	Secreted protein g
c 299	47.2	4.7	2420	22	AAL26460	Human breast cance
c 300	47.2	4.7	2421	22	AAL26449	Human breast cance

ALIGNMENTS

```
RESULT 1
AAI64291
ID   AAI64291 standard; DNA; 128139 BP.
XX
AC   AAI64291;
XX
DT   22-APR-2002  (first entry)
XX
DE   RRV genome nucleotide sequence.
XX
KW   RRV; rhesus rhadinovirus; japanese macaque virus; multiple sclerosis;
KW   JMHV; cytostatic; antiasthmatic; antiallergic; dermatological;
KW   vulnerary; gene therapy; leucopenia; thrombocytopaenia;
KW   inflammatory disease; asthma; allergy; dermatitis; virus; ds.
XX
OS   Macaca mulatta rhadinovirus 17577.
XX
FH   Key             Location/Qualifiers
FT   CDS             1353..2674
FT                   /*tag= a
FT                   /product= "RRV R1"
FT   CDS             complement (2692..3258)
FT                   /*tag= b
FT                   /product= "dihydrofolate reductase"
FT                   /label= RRV_ORF2
FT                   /note= "has similarity to Kaposi's sarcoma-associated
FT                         virus (KSHV) open reading frame (ORF) 2"
FT   CDS             3676..5613
FT                   /*tag= c
FT                   /product= "complement binding protein"
FT                   /label= RRV_ORF4
FT                   /note= "has similarity to KSHV ORF4"
FT   CDS             6045..9443
FT                   /*tag= d
FT                   /product= "ssDNA binding protein"
FT                   /label= RRV_ORF6
FT                   /note= "has similarity to KSHV ORF6"
FT   CDS             9468..11528
FT                   /*tag= e
FT                   /product= "transport protein"
FT                   /label= RRV_ORF7
FT                   /note= "has similarity to KSHV ORF7"
FT   CDS             11515..14004
FT                   /*tag= f
FT                   /product= "glycoprotein B"
FT                   /label= RRV_ORF8
FT                   /note= "has similarity to KSHV ORF8"
FT   CDS             14122..17166
FT                   /*tag= g
FT                   /product= "DNA polymerase protein"
FT                   /label= RRV_ORF9
FT                   /note= "has similarity to KSHV ORF9"
FT   CDS             17261..18511
FT                   /*tag= h
FT                   /label= RRV_ORF10
FT                   /note= "has similarity to KSHV ORF10"
FT   CDS             18520..19749
FT                   /*tag= i
FT                   /label= RRV_ORF11
FT                   /note= "has similarity to KSHV ORF11"
FT   CDS             complement (19921..20544)
FT                   /*tag= j
FT                   /product= "RRV R2"
FT                   /label= RRV_ORF
FT                   /note= "has similarity to KSHV interleukin (IL)-6 gene"
FT   CDS             complement (20777..21778)
FT                   /*tag= k
FT                   /product= "thymidylate synthase"
```

```
FT                   /label= RRV_ORF70
FT                   /note= "has similarity to KSHV ORF70"
FT   CDS             complement (22245..22592)
FT                   /*tag=  l
FT                   /product= "RRV R3"
FT                   /note= "has similarity to KSHV K4 viral MIP gene"
FT   CDS             26846..27409
FT                   /*tag=  m
FT                   /product= "Bcl2-homologue"
FT                   /label= RRV_ORF16
FT                   /note= "has similarity to KSHV ORF16"
FT   CDS             complememt (27515..29125)
FT                   /*tag=  n
FT                   /label= RRV_ORF17
FT                   /note= "has similarity to KSHV ORF17"
FT   CDS             28998..29897
FT                   /*tag=  o
FT                   /label= RRV_ORF18
FT                   /note= "has similarity to KSHV ORF18"
FT   CDS             complement (29905..31548)
FT                   /*tag=  p
FT                   /product= "tegument protein
FT                   /label= RRV_ORF19
FT                   /note= "has similarity to KSHV ORF19"
FT   CDS             complement (31043..32095)
FT                   /*tag=  q
FT                   /label= RRV_ORF20
FT                   /note= "has similarity to KSHV ORF20"
FT   CDS             32094..33767
FT                   /*tag=  r
FT                   /product= "thymidine kinase"
FT                   /label= RRV_ORF21
FT                   /note= "has similarity to KSHV ORF21"
FT   CDS             33754..35868
FT                   /*tag=  s
FT                   /product= "glycoprotein H"
FT                   /label= RRV_ORF22
FT                   /note= "has similarity to KSHV ORF22"
FT   CDS             complement (35865..37073)
FT                   /*tag=  t
FT                   /label= RRV_ORF23
FT                   /note= "has similarity to KSHV ORF23"
FT   CDS             complement (37123..39321)
FT                   /*tag=  u
FT                   /label= RRV_ORF24
FT                   /note= "has similarity to KSHV ORF24"
FT   CDS             39323..43459
FT                   /*tag=  v
FT                   /product= "major capsid protein"
FT                   /label= RRV_ORF25
FT                   /note= "has similarity to KSHV ORF25"
FT   CDS             43491..44408
FT                   /*tag=  w
FT                   /product= "capsid protein"
FT                   /label= RRV_ORF26
FT                   /note= "has similarity to KSHV ORF26"
FT   CDS             44433..45242
FT                   /*tag=  x
FT                   /label= RRV_ORF27
FT                   /note= "has similarity to KSHV ORF27"
FT   CDS             45408..45683
FT                   /*tag=  y
FT                   /label= RRV_ORF28
FT                   /note= "has similarity to KSHV ORF28"
FT   CDS             complement (45733..46779)
FT                   /*tag=  z
FT                   /label= RRV_ORF29b
FT                   /note= "has similarity to KSHV ORF29b"
FT   CDS             46905..47135
FT                   /*tag=  aa
FT                   /label= RRV_ORF30
FT                   /note= "has similarity to KSHV ORF30"
FT   CDS             47093..47746
FT                   /*tag=  ab
FT                   /label= RRV_ORF31
FT                   /note= "has similarity to KSHV ORF31"
FT   CDS             47683..49077
FT                   /*tag=  ac
FT                   /label= RRV_ORF32
FT                   /note= "has similarity to KSHV ORF32"
FT   CDS             49049..50059
FT                   /*tag=  ad
FT                   /label= RRV_ORF33
FT                   /note= "has similarity to KSHV ORF33"
FT   CDS             complement (49977..50960)
FT                   /*tag=  ae
FT                   /label= RRV_ORF29a
FT                   /note= "has similarity to KSHV ORF29a"
FT   CDS             50959..51942
FT                   /*tag=  af
FT                   /label= RRV_ORF34
FT                   /note= "has similarity to KSHV ORF34"
FT   CDS             51923..52372
FT                   /*tag=  ag
FT                   /label= RRV_ORF35
FT                   /note= "has similarity to KSHV ORF35"
FT   CDS             52278..53585
FT                   /*tag=  ah
FT                   /product= "kinase"
FT                   /label= RRV_ORF36
FT                   /note= "has similarity to KSHV ORF36"
FT   CDS             53566..55008
FT                   /*tag=  ai
FT                   /product= "alkaline exonuclease"
FT                   /label= RRV_ORF37
FT                   /note= "has similarity to KSHV ORF37"
FT   CDS             54963..55172
FT                   /*tag=  aj
FT                   /label= RRV_ORF38
FT                   /note= "has similarity to KSHV ORF38"
FT   CDS             complement (55255..56391)
FT                   /*tag=  ak
FT                   /product= "glycoprotein M"
FT                   /label= RRV_ORF39
FT                   /note= "has similarity to KSHV ORF39"
FT   CDS             56526..57932
FT                   /*tag=  al
FT                   /product= "helicase/primase"
FT                   /label= RRV_ORF40
FT                   /note= "has similarity to KSHV ORF40"
FT   CDS             57917..58528
FT                   /*tag=  am
FT                   /product= "helicase/primase"
FT                   /label= RRV_ORF41
FT                   /note= "has similarity to KSHV ORF41"
FT   CDS             complement (58525..59343)
FT                   /*tag=  an
FT                   /label= RRV_ORF42
FT                   /note= "has similarity to KSHV ORF42"
FT   CDS             complement (59297..61027)
FT                   /*tag=  ao
FT                   /product= "capsid protein"
FT                   /label= RRV_ORF43
FT                   /note= "has similarity to KSHV ORF43"
```

```
  Query Match             100.0%;  Score 1001;  DB 24;  Length 128139;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-179;
  Matches 1001;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CGCCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127139 CGCCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTG 127198

Qy         61 GTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127199 GTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGG 127258
```

```
Qy        121 GGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127259 GGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGAC 127318

Qy        181 AGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127319 AGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGC 127378

Qy        241 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127379 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 127438

Qy        301 AGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127439 AGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 127498

Qy        361 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127499 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 127558

Qy        421 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127559 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 127618

Qy        481 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127619 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 127678

Qy        541 CTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127679 CTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGA 127738

Qy        601 TGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127739 TGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 127798

Qy        661 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127799 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 127858

Qy        721 TGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127859 TGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 127918

Qy        781 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127919 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGC 127978

Qy        841 TTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127979 TTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAG 128038

Qy        901 GGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACCCGGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     128039 GGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACCCGGG 128098

Qy        961 CCTCTCCCCCGCCCGGACCGCCGGGCACCCGGCCCAGATCT 1001
              |||||||||||||||||||||||||||||||||||||||||
Db     128099 CCTCTCCCCCGCCCGGACCGCCGGGCACCCGGCCCAGATCT 128139
```

RESULT 2
AAC64754
ID AAC64754 standard; DNA; 133719 BP.
XX
AC AAC64754;
XX
DT 28-FEB-2001 (first entry)
XX
DE Macaca mulatta rhadinovirus 17577 (RRV) genome sequence SEQ ID NO:1.
XX
KW Macaca mulatta rhadinovirus 17577; RRV; rhesus macaque rhadinovirus;
KW genome; Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6;
KW IL-6; macrophage inflammatory protein; MIP; diagnosis; vaccine;
KW cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma;
KW lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy;
KW splenomegaly; hypergammaglobinulinaemia; autoimmune haemolytic anaemia;
KW ds.
XX
OS Macaca mulatta rhadinovirus 17577.
XX
PN WO200028040-A2.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26260.
XX
PR 06-NOV-1998; 98US-0107507.
PR 20-NOV-1998; 98US-0109409.
XX
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI Wong SW, Axthelm MK, Searles RP;
XX
DR WPI; 2000-376552/32.
XX
PT New rhesus rhadino virus for producing non-human primate model useful
PT for testing potential treatments and efficacy of the candidate vaccine
PT for conditions associated with RRV infection -
XX
PS Claim 2; Page 83-122; 141pp; English.
XX
CC The present invention describes a novel rhesus macaque rhadinovirus
CC called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the
CC RRV genome sequence, and AAB53123 to AAB53204 represent the proteins
CC encoded by the genome sequence. The present invention also specifically
CC claims the individual open reading frame (ORF) nucleotide sequences from
CC the genome which encode the individual proteins, but these sequences are
CC not given. A non-human animal infected with RRV can be used for testing
CC the efficacy of drug in the treatment of condition associated with
CC infection with RRV such as Kaposi's sarcoma, lymphoproliferative
CC disorders, B-cell hyperplasia, lymphadenopathy, splenomegaly,
CC hypergammaglobinulinaemia or autoimmune haemolytic anaemia, by
CC administering the drug to a immuno-compromised non-human primate
CC preferably Rhesus macaque monkey obtained by as a result of infection
CC by Simian Immunodeficiency Virus (SIV). RRV is useful for producing
CC non-human primate model for testing potential treatments for conditions
CC associated with RRV infection. It is also useful for testing the
CC efficacy of the candidate vaccine against RRV infection or conditions
CC associated with its infection by administering the vaccine to the
CC subject capable of infection with RRV, inoculating the subject with RRV
CC and observing the effect of vaccine. AAC64755 to AAC64765 and AAB53205
CC to AAB53213 represent sequence used in the exemplification of the
CC present invention.
XX
SQ Sequence 133719 BP; 32746 A; 35648 C; 34521 G; 30804 T; 0 other;

```
  Query Match             100.0%;  Score 1001;  DB 21;  Length 133719;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-179;
  Matches 1001;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CGCCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     132719 CGCCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTG 132778

Qy         61 GTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     132779 GTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGG 132838

Qy        121 GGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     132839 GGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGAC 132898

Qy        181 AGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     132899 AGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGC 132958

Qy        241 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     132959 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 133018

Qy        301 AGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133019 AGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 133078

Qy        361 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133079 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 133138

Qy        421 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133139 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 133198

Qy        481 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133199 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 133258

Qy        541 CTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133259 CTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGA 133318

Qy        601 TGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133319 TGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 133378

Qy        661 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133379 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 133438

Qy        721 TGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133439 TGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 133498

Qy        781 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133499 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGC 133558

Qy        841 TTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133559 TTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAG 133618

Qy        901 GGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACCCGGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133619 GGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACCCGGG 133678

Qy        961 CCTCTCCCCCGCCCGGACCGCCGGGCACCCGGCCCAGATCT 1001
              |||||||||||||||||||||||||||||||||||||||||
Db     133679 CCTCTCCCCCGCCCGGACCGCCGGGCACCCGGCCCAGATCT 133719


RESULT 3
ABA41262
ID   ABA41262 standard; DNA; 576 BP.
XX
AC   ABA41262;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Probe #19728 for gene expression analysis in human heart cell sample.
XX
KW   Human; gene expression; heart; microarray; vascular system; probe;
KW   cardiovascular disease; hypertension; cardiac arrhythmia;
KW   congenital heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157274-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00666.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488899/53.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   hearts -
XX
PS   Claim 4; SEQ ID No 19728; 530pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes for
CC   measuring human gene expression in a sample derived from human heart. The
CC   present sequence is one such probe. The probes may be used for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from the human heart via microarrays. By measuring gene expression, the
CC   probes are useful for predicting, diagnosing, grading, staging,
CC   monitoring and prognosing diseases of the human heart and vascular system
CC   e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC   congenital heart disease.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 576 BP; 7 A; 60 C; 219 G; 290 T; 0 other;

  Query Match             19.3%;  Score 193.4;  DB 22;  Length 576;
  Best Local Similarity   59.0%;  Pred. No. 9.3e-28;
  Matches  332;  Conservative    0;  Mismatches  231;  Indels    0;  Gaps    0;

Qy        303 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 362
              || || ||| |  |       |   || || ||| | ||| |  |||||| |  | |
Db          1 GCTTGGTTGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTTGGCTGGTT 60

Qy        363 GTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 422
              |   || || | ||||| | |   || || | | | |  ||||||||||||| ||| || |
Db         61 GCTTGGTTGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTGCTT 120

Qy        423 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 482
              |||||||||| ||| || ||  || ||| | | | |  |||||| ||| | | | || | |
Db        121 GCTTGCTTGCTTGCTTGCTTGGCTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGCTT 180

Qy        483 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 542
              |||||||||| | ||| || |  ||| ||||| | | || |||||  ||| | ||| |  |
Db        181 GCTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTT 240

Qy        543 GCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATG 602
              |||||||  |  |   |   || || |||||||  |  |  |   |   |  ||  || |
Db        241 GCTTGCTTCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTTGGTTGGTT 300

Qy        603 GCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 662
              | ||| | |   || |||| || |||| |   || || | || || | | ||   || ||||
Db        301 GGTTGCTTGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGTTTGGTTGCTT 360

Qy        663 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTG 722
              || || | |   || || | |  || | |   || |  | |  |||| |    | ||||
Db        361 GGTTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTTGTTGCTT 420

Qy        723 GGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 782
              |  ||  |      || || | || || | |   || || | || || | |   || || |
Db        421 GCTTGGTTCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 480
```

```
Qy        783 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTT 842
              | || | |   || || | || |||| |   || || ||| |   ||| ||| ||| ||
Db        481 CGTTGATTGCTTGGTTGGTTGGTTGCTTGGTTGGTTGGCTGGTTCATTGGTTGGTTGGTT 540

Qy        843 GCTAGTGGGGCCGCTTGCCTGCT 865
              ||| |   |   | |||  || |
Db        541 GCTTGGTTGTTTGGTTGGTTGGT 563


RESULT 4
AAI57480
ID   AAI57480 standard; DNA; 576 BP.
XX
AC   AAI57480;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #26166 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 26166; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 576 BP; 7 A; 60 C; 219 G; 290 T; 0 other;

  Query Match             19.3%;  Score 193.4;  DB 22;  Length 576;
  Best Local Similarity   59.0%;  Pred. No. 9.3e-28;
  Matches  332;  Conservative    0;  Mismatches  231;  Indels    0;  Gaps    0;

Qy        303 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 362
              || || ||| |  |      |   || || ||| | ||| |  |||||| |  |  |
Db          1 GCTTGGTTGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTTGGCTGGTT 60

Qy        363 GTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 422
              |   || || | ||||| | |   || || | | | |  ||||||||||| ||| || |
Db         61 GCTTGGTTGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTGCTT 120

Qy        423 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 482
              |||||||||| ||| || ||  || ||| | | | |  |||||| ||| | | | || |
Db        121 GCTTGCTTGCTTGCTTGCTTGGCTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGCTT 180

Qy        483 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 542
              ||||||||| | ||| || |  ||| |||||  | | || |||||  ||| | ||| |  |
Db        181 GCTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTT 240

Qy        543 GCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATG 602
              |||||||  |  |   |   || || |||||||  |  |  |   |   |  ||  || |
Db        241 GCTTGCTTCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTTGGTTGGTT 300

Qy        603 GCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 662
              | ||| | |   || |||| || |||| |   || || | || || | ||  || ||||
Db        301 GGTTGCTTGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGTTTGGTTGCTT 360

Qy        663 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTG 722
              || || | |   || || | |  || | |   || |  | |  |||| |    | ||||
Db        361 GGTTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTTGTTGCTT 420

Qy        723 GGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 782
              |  || |     || || | || || | |   || || | || || | |   || || |
Db        421 GCTTGGTTCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 480

Qy        783 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTT 842
              | || | |   || || | || |||| |   || || ||| |   ||| ||| ||| ||
Db        481 CGTTGATTGCTTGGTTGGTTGGTTGCTTGGTTGGTTGGCTGGTTCATTGGTTGGTTGGTT 540

Qy        843 GCTAGTGGGGCCGCTTGCCTGCT 865
              ||| |   |   | |||  || |
Db        541 GCTTGGTTGTTTGGTTGGTTGGT 563


RESULT 5
ABA31139
ID   ABA31139 standard; DNA; 555 BP.
XX
AC   ABA31139;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Probe #9605 for gene expression analysis in human heart cell sample.
XX
KW   Human; gene expression; heart; microarray; vascular system; probe;
KW   cardiovascular disease; hypertension; cardiac arrhythmia;
KW   congenital heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157274-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00666.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488899/53.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   hearts -
XX
PS   Claim 1; SEQ ID No 9605; 530pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes for
CC   measuring human gene expression in a sample derived from human heart. The
```

```
CC   present sequence is one such probe. The probes may be used for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from the human heart via microarrays. By measuring gene expression, the
CC   probes are useful for predicting, diagnosing, grading, staging,
CC   monitoring and prognosing diseases of the human heart and vascular system
CC   e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC   congenital heart disease.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 555 BP; 18 A; 59 C; 201 G; 277 T; 0 other;

  Query Match             17.4%;  Score 174.6;  DB 22;  Length 555;
  Best Local Similarity   58.9%;  Pred. No. 3.2e-24;
  Matches  300;  Conservative    0;  Mismatches  209;  Indels    0;  Gaps    0;

Qy        337 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGC 396
Db         46 TTGGTTGACTGGTTGCTTGGTTGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGC 105

Qy        397 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 456
Db        106 TTGGTTGGCTGGTTGCTTGGTTGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGC 165

Qy        457 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 516
Db        166 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGGCTGGTTGGTTGGTTGGTTGCTTGG 225

Qy        517 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGC 576
Db        226 TTGGTTGGTTGCTTGCTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCA 285

Qy        577 TAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGG 636
Db        286 TTGGTTGCTTGGTTGCTTGCTTCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGG 345

Qy        637 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 696
Db        346 TTGGTTGGTTGGTTGGTTGCTTGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGG 405

Qy        697 CTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGG 756
Db        406 TTGTTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGC 465

Qy        757 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 816
Db        466 TTGCTTTGTTGCTTGCTTGGTTCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGG 525

Qy        817 CTGCCTGCTGGCTTGCTTGCTTGCTTGCT 845
Db        526 TTGGTTGGTTGGTTCGTTGATTGCTTGGT 554


RESULT 6
AAI44102
ID   AAI44102 standard; DNA; 555 BP.
XX
AC   AAI44102;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #12788 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 12788; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 555 BP; 18 A; 59 C; 201 G; 277 T; 0 other;

  Query Match             17.4%;  Score 174.6;  DB 22;  Length 555;
  Best Local Similarity   58.9%;  Pred. No. 3.2e-24;
  Matches  300;  Conservative    0;  Mismatches  209;  Indels    0;  Gaps    0;

Qy        337 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGC 396
Db         46 TTGGTTGACTGGTTGCTTGGTTGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGC 105

Qy        397 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 456
Db        106 TTGGTTGGCTGGTTGCTTGGTTGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGC 165

Qy        457 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 516
Db        166 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGGCTGGTTGGTTGGTTGGTTGCTTGG 225

Qy        517 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGC 576
Db        226 TTGGTTGGTTGCTTGCTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCA 285

Qy        577 TAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGG 636
Db        286 TTGGTTGCTTGGTTGCTTGCTTCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGG 345

Qy        637 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 696
Db        346 TTGGTTGGTTGGTTGGTTGCTTGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGG 405

Qy        697 CTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGG 756
Db        406 TTGTTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGC 465

Qy        757 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 816
Db        466 TTGCTTTGTTGCTTGCTTGGTTCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGG 525

Qy        817 CTGCCTGCTGGCTTGCTTGCTTGCTTGCT 845
Db        526 TTGGTTGGTTGGTTCGTTGATTGCTTGGT 554


RESULT 7
AAC64754/c
ID   AAC64754 standard; DNA; 133719 BP.
```

```
XX
AC   AAC64754;
XX
DT   28-FEB-2001  (first entry)
XX
DE   Macaca mulatta rhadinovirus 17577 (RRV) genome sequence SEQ ID NO:1.
XX
KW   Macaca mulatta rhadinovirus 17577; RRV; rhesus macaque rhadinovirus;
KW   genome; Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6;
KW   IL-6; macrophage inflammatory protein; MIP; diagnosis; vaccine;
KW   cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma;
KW   lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy;
KW   splenomegaly; hypergammaglobinulinaemia; autoimmune haemolytic anaemia;
KW   ds.
XX
OS   Macaca mulatta rhadinovirus 17577.
XX
PN   WO200028040-A2.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26260.
XX
PR   06-NOV-1998;   98US-0107507.
PR   20-NOV-1998;   98US-0109409.
XX
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Wong SW,  Axthelm MK,  Searles RP;
XX
DR   WPI; 2000-376552/32.
XX
PT   New rhesus rhadino virus for producing non-human primate model useful
PT   for testing potential treatments and efficacy of the candidate vaccine
PT   for conditions associated with RRV infection  -
XX
PS   Claim 2; Page 83-122; 141pp; English.
XX
CC   The present invention describes a novel rhesus macaque rhadinovirus
CC   called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the
CC   RRV genome sequence, and AAB53123 to AAB53204 represent the proteins
CC   encoded by the genome sequence. The present invention also specifically
CC   claims the individual open reading frame (ORF) nucleotide sequences from
CC   the genome which encode the individual proteins, but these sequences are
CC   not given. A non-human animal infected with RRV can be used for testing
CC   the efficacy of drug in the treatment of condition associated with
CC   infection with RRV such as Kaposi's sarcoma, lymphoproliferative
CC   disorders, B-cell hyperplasia, lymphadenopathy, splenomegaly,
CC   hypergammaglobinulinaemia or autoimmune haemolytic anaemia, by
CC   administering the drug to a immuno-compromised non-human primate
CC   preferably Rhesus macaque monkey obtained by as a result of infection
CC   by Simian Immunodeficiency Virus (SIV). RRV is useful for producing
CC   non-human primate model for testing potential treatments for conditions
CC   associated with RRV infection. It is also useful for testing the
CC   efficacy of the candidate vaccine against RRV infection or conditions
CC   associated with its infection by administering the vaccine to the
CC   subject capable of infection with RRV, inoculating the subject with RRV
CC   and observing the effect of vaccine. AAC64755 to AAC64765 and AAB53205
CC   to AAB53213 represent sequence used in the exemplification of the
CC   present invention.
XX
SQ   Sequence 133719 BP; 32746 A; 35648 C; 34521 G; 30804 T; 0 other;

  Query Match           13.4%;  Score 134.2;  DB 21;  Length 133719;
  Best Local Similarity   51.6%;  Pred. No. 1.4e-16;
  Matches  333;  Conservative    0;  Mismatches  308;  Indels    4;  Gaps    1;

Qy        221 CTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGC 280
              | ||   |||  | || |||  ||   |  |   |  || | ||  ||  ||  || |||
Db        748 CAGCAGCCCTAGTAGCAGGCAAGCGGCCCCACTAGCAAGCAAGCAAGCAAGCCAGCAGGC 689

Qy        281 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGC 340
                | | ||    |  ||| || ||| || ||     |  ||| || ||| || ||   |

Db        688 AGCCCTACTAGCAGCCCAGCAGCCCTACTAGC----AGCCCAGCAGCCCTACTAGCAGCC 633

Qy        341 TGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGC 400
                ||   | | ||    |  ||| ||  || || ||   | || ||  || || | | |
Db        632 CAGCAGCCCTACTAGCAGCCCAGGAGCCCTACTAGCAGCCCAGCAGCCCTACTAGCAGCC 573

Qy        401 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 460
              | |    | | || || | || ||   | | || || | || ||   | | || || | |
Db        572 CAGGAGCCCTACTAGCAGCCCAGCAGCCCTACTAGCAGCCCAGCAGCCCTACTAGCAGCC 513

Qy        461 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 520
              | ||   | | || || | || ||   | | ||  |  |||  |  ||  ||   | | |
Db        512 CAGCAGCCCTACTAGCAGCCCAGCAGCCCTACTATCAAGCCATCAGGCAGGCGTACCGTC 453

Qy        521 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAG 580
              |   | ||  ||  || | ||| ||  |       |  || ||| | ||  ||  || | |
Db        452 CCCTTAGCAAGCAAGCAGCCCTACTGTCCCCTCAGCAAGCAGGCCAGCGAGCAAGCAAGC 393

Qy        581 GGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 640
              || | |   ||  || | ||| |  || | |   || || | || || | | ||  ||
Db        392 AGGCCAGCGAGCAAGCAAGCAGGCCAGCGAGCAAGCAAGCAGGCCAGCGAGCAAGCAAGC 333

Qy        641 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCC 700
              || | || || | |   || || | || || | |   || || | || || | |    |
Db        332 AGGCCAGCGAGCAAGCAAGCAGGCCAGCGAGCAAGCAAGCAGGCCAGCGAGCAAGCAAGC 273

Qy        701 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 760
              || | || || | |   || || | || || ||  || || |   || || ||  || |
Db        272 AGGCCAGCGAGCAAGCAAGCAGGCCAGCGAGCAGCCCTACTAGCAAGCCAGCAGCCCTAC 213

Qy        761 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 820
              ||  |    || | |   ||  |   || || ||| || |||  |    || | |   ||
Db        212 TGTCCCCTCAGCAAGCAGGCCAGCGAGCAAGCAGGCCTACTGTCCCCTCAGCAAGCAGGC 153

Qy        821 CTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCT 865
              || ||| |      ||  ||  || || | |   ||  |||| ||
Db        152 CTACTGTCCCCTCAGCAAGCAGGCCAGCGAGCAAGCAGGCCTACT 108


RESULT 8
AAV73802/c
ID   AAV73802 standard; DNA; 35100 BP.
XX
AC   AAV73802;
XX
DT   25-FEB-1999  (first entry)
XX
DE   KSHV LUR DNA (nucleotides 1-35,100).
XX
KW   Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHFR; Bcl-2;
KW   dihydrofolate reductase; LUR; long unique region; vaccine; prophylaxis;
KW   diagnosis; treatment; HHV8; complement binding protein; v-CBP; SSBP;
KW   ssDNA binding protein; transport protein; glycoprotein B; pol; vIL-6;
KW   DNA polymerase; viral interleukin-6; BHV4-IE1 I; thymidylate synthase;
KW   vMIP-II; BHV4-IE1 II; vMIP-I; capsid protein I; tegument protein I; ds.
XX
OS   Kaposi's sarcoma-associated herpesvirus.
XX
PN   US5849564-A.
XX
PD   15-DEC-1998.
XX
PF   29-NOV-1996;   96US-0770379.
XX
PR   29-NOV-1996;   96US-0770379.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Bohenzky RA,  Chang Y,  Edelman IS,  Moore PS,  Russo JJ;
XX
DR   WPI; 1999-069741/06.
XX
PT   Kaposi's sarcoma-associated herpes virus nucleic acid - encodes
```

```
PT   di:hydro:folate reductase and is useful for treatment, prophylaxis
PT   or diagnosis of Kaposi's sarcoma
XX
PS   Disclosure; Column 67-96; 109pp; English.
XX
CC   This sequence is a fragment of the Kaposi's sarcoma-associated
CC   herpesvirus (KSHV) LUR (long unique region). This fragment contains
CC   coding regions for K1, ORF4 which encodes the complement binding protein
CC   v-CBP, ORF6 which encodes a ssDNA binding protein (SSBP), ORF7 which
CC   encodes a transport protein, ORF8 which encodes glycoprotein B, ORF9
CC   which encodes DNA polymerase (pol), ORF10, OFR11, K2 which encodes viral
CC   interleukin-6 (vIL-6), ORF02 which encodes dihydrofolate reductase
CC   (DHFR), K3 which encodes BHV4-IE1 I, ORF70 which encodes Thymidylate
CC   synthase, K4 which encodes vMIP-II, K5 which encodes BHV4-IE1 II, K6
CC   which encodes vMIP-I, K7, ORF16 which encodes Bcl-2, ORF17 which encodes
CC   capsid protein I, ORF18 and ORF19 which encodes Tegument protein I.
CC   KSHV is a new human Herpesvirus (HHV8) believed to cause Kaposi's sarcoma
CC   (KS) which is the most common form of neoplasm occurring in persons with
CC   acquired immune deficiency syndrome (AIDS). The DHFR protein is useful
CC   for vaccination, prophylaxis, diagnosis and treatment of a subject with
CC   Kaposi's sarcoma and for detecting expression of a DNA virus associated
CC   with Kaposi's sarcoma in a cell.
XX
SQ   Sequence 35100 BP; 8703 A; 9395 C; 8921 G; 8081 T; 0 other;

  Query Match             12.7%;  Score 127.6;  DB 20;  Length 35100;
  Best Local Similarity   66.9%;  Pred. No. 2.4e-15;
  Matches  198;  Conservative    0;  Mismatches   94;  Indels    4;  Gaps    1;

Qy        595 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 654
              | ||  |||  || |    ||||| ||||| ||||    ||||| ||||| ||||    |
Db      24611 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24552

Qy        655 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 714
              |||| ||||| ||||    ||||| ||||| ||||    |||||||||| ||||    |
Db      24551 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24492

Qy        715 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 774
              |||| ||||| ||||    |||||||||| ||||    ||||| ||||| ||||    |
Db      24491 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24432

Qy        775 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT----GCCTGCTGGCTT 830
              |||| ||||| ||||    ||||| ||||| ||||    |||||    |  ||||| |
Db      24431 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24372

Qy        831 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTG 886
              |  | || |  ||||   |||||  || |  |||| |  ||||| ||| |  ||||
Db      24371 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTG 24316


RESULT 9
AAV19941/c
ID   AAV19941 standard; DNA; 137507 BP.
XX
AC   AAV19941;
XX
DT   03-AUG-1998  (first entry)
XX
DE   KSHV long unique coding region and terminal repeat.
XX
KW   KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II;
KW   interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;
KW   complement-binding protein; glycoprotein; capsid protein IV; infection;
KW   immediate early protein; Kaposi's sarcoma; protective vaccine; lymphoma;
KW   lymphoproliferative disease; leukaemia; splenomegaly; mycosis fungoides;
KW   HIV immune status; anti-inflammatory agent; therapy; ds.
XX
OS   Kaposi's sarcoma-associated herpes virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1142..2794
FT                   /*tag= a
FT                   /product= complement-binding protein
FT   CDS             8699..11236
FT                   /*tag= b
FT                   /product= glycoprotein B
FT   CDS             complement (17261..17875)
FT                   /*tag= c
FT                   /product= interleukin 6
FT   CDS             complement (21548..21832)
FT                   /*tag= d
FT                   /product= macrophage inflammatory protein II
FT   CDS             complement (27137..27424)
FT                   /*tag= e
FT                   /product= interferon regulatory factor 1
FT   CDS             28661..29741
FT                   /*tag= f
FT                   /product= protein T1.1
FT   CDS             complement (58976..60175)
FT                   /*tag= g
FT                   /product= glycoprotein M
FT   CDS             complement (69412..69915)
FT                   /*tag= h
FT                   /product= glycoprotein L
FT   CDS             complement (88410..88910)
FT                   /*tag= i
FT                   /product= interferon regulatory factor 2
FT   CDS             89600..90541
FT                   /*tag= j
FT                   /product= interferon regulatory factor 3
FT   CDS             90173..90643
FT                   /*tag= k
FT                   /product= glycoprotein X
FT   CDS             complement (93636..94127)
FT                   /*tag= l
FT                   /product= interferon regulatory factor 4
FT   CDS             complement (111931..112443)
FT                   /*tag= m
FT                   /product= capsid protein IV
FT   CDS             complement (123808..127296)
FT                   /*tag= n
FT                   /product= immediate early protein
XX
PN   WO9804576-A1.
XX
PD   05-FEB-1998.
XX
PF   22-JUL-1997;   97WO-US13346.
XX
PR   29-NOV-1996;   96US-0757669.
PR   25-JUL-1996;   96US-0686243.
PR   25-JUL-1996;   96US-0686349.
PR   25-JUL-1996;   96US-0686350.
PR   25-JUL-1996;   96US-0687253.
PR   25-JUL-1996;   96US-0688814.
PR   05-SEP-1996;   96US-0708678.
PR   10-OCT-1996;   96US-0728323.
PR   13-NOV-1996;   96US-0747887.
PR   13-NOV-1996;   96US-0748640.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Bohenzky RA,  Chang Y,  Edelman IS,  Moore PS,  Russo JJ;
XX
DR   WPI; 1998-130615/12.
XX
PT   New nucleic acid encoding Kaposi's sarcoma associated herpes virus
PT   proteins - useful for, e.g. detecting levels of HHV8 in, and
PT   preparation of vaccines for treatment of, HIV patients
XX
PS   Example 2; Page 135-203; 230pp; English.
XX
CC   This sequence represents the long unique region and terminal repeat of
CC   the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known
CC   as human herpes virus 8 (HHV8). This sequence contains the DNAs of the
CC   invention which encode KSHV polypeptides selected from: (a) viral
```

```
CC   macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);
CC   (c) viral IRF 1; (d) complement-binding protein; glycoproteins B, M or L;
CC   (d) capsid protein IV encoded by ORF65; and (e) immediate early protein
CC   encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded
CC   by it, and antibodies (Ab) specific for the proteins are useful for
CC   detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body
CC   fluids or tissue samples. HHV8 infections can be treated with antisense
CC   or triplex forming molecules or agents that bind specifically to the
CC   protein. Ab may be used for prophylaxis or treatment of HHV8 infection,
CC   while the protein can be used in protective vaccines. Ab may also be used
CC   to differentiate between lymphomas, and HHV8 may be implicated in many
CC   other lymphoproliferative diseases such as lymphomas, leukaemia,
CC   splenomegaly and mycosis fungoides. Cells and animals containing the
CC   nucleic acid are useful for drug screening. HHV8-derived peptides can be
CC   used as targets for antiviral drugs, e.g. dihydrofolate reductase gene
CC   can be inhibited with methotrexate. These can also be used to determine
CC   the immune status of a patient infected with HIV. HHV8 derived protein
CC   viral MIP III may be used as an anti-inflammatory agent for,
CC   e.g. treating rheumatoid arthritis. This sequence is stated as containing
CC   81 open reading frames.
XX
SQ   Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T; 0 other;

  Query Match             12.7%;  Score 127.6;  DB 19;  Length 137507;
  Best Local Similarity   66.9%;  Pred. No. 2.5e-15;
  Matches  198;  Conservative    0;  Mismatches   94;  Indels    4;  Gaps    1;

Qy        595 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 654
              | ||  |||  || |    ||||| ||||| ||||    ||||| ||||| ||||    |
Db      24611 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24552

Qy        655 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 714
              |||| |||.|| ||||    ||||| ||||| ||||    |||||||||||| ||||    |
Db      24551 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24492

Qy        715 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 774
              |||| ||||| ||||    |||||||||||| ||||    ||||| ||||| ||||    |
Db      24491 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24432

Qy        775 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT----GCCTGCTGGCTT 830
              |||| ||||| ||||    ||||| ||||| ||||    |||||    |  ||||| |
Db      24431 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24372

Qy        831 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTG 886
              |  | || |  |||   |||||  || |  |||| |  ||||| ||| |  |||
Db      24371 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTG 24316


RESULT 10
AAI64291/c
ID   AAI64291 standard; DNA; 128139 BP.
XX
AC   AAI64291;
XX
DT   22-APR-2002  (first entry)
XX
DE   RRV genome nucleotide sequence.
XX
KW   RRV; rhesus rhadinovirus; japanese macaque virus; multiple sclerosis;
KW   JMHV; cytostatic; antiasthmatic; antiallergic; dermatological;
KW   vulnerary; gene therapy; leucopenia; thrombocytopaenia;
KW   inflammatory disease; asthma; allergy; dermatitis; virus; ds.
XX
OS   Macaca mulatta rhadinovirus 17577.
XX
FH   Key             Location/Qualifiers
FT   CDS             1353..2674
FT                   /*tag= a
FT                   /product= "RRV R1"
FT   CDS             complement (2692..3258)
FT                   /*tag= b
FT                   /product= "dihydrofolate reductase"
FT                   /label= RRV_ORF2
FT                   /note= "has similarity to Kaposi's sarcoma-associated
FT                   virus (KSHV) open reading frame (ORF) 2"
FT   CDS             3676..5613
FT                   /*tag= c
FT                   /product= "complement binding protein"
FT                   /label= RRV_ORF4
FT                   /note= "has similarity to KSHV ORF4"
FT   CDS             6045..9443
FT                   /*tag= d
FT                   /product= "ssDNA binding protein"
FT                   /label= RRV_ORF6
FT                   /note= "has similarity to KSHV ORF6"
FT   CDS             9468..11528
FT                   /*tag= e
FT                   /product= "transport protein"
FT                   /label= RRV_ORF7
FT                   /note= "has similarity to KSHV ORF7"
FT   CDS             11515..14004
FT                   /*tag= f
FT                   /product= "glycoprotein B"
FT                   /label= RRV_ORF8
FT                   /note= "has similarity to KSHV ORF8"
FT   CDS             14122..17166
FT                   /*tag= g
FT                   /product= "DNA polymerase protein"
FT                   /label= RRV_ORF9
FT                   /note= "has similarity to KSHV ORF9"
FT   CDS             17261..18511
FT                   /*tag= h
FT                   /label= RRV_ORF10
FT                   /note= "has similarity to KSHV ORF10"
FT   CDS             18520..19749
FT                   /*tag= i
FT                   /label= RRV_ORF11
FT                   /note= "has similarity to KSHV ORF11"
FT   CDS             complement (19921..20544)
FT                   /*tag= j
FT                   /product= "RRV R2"
FT                   /label= RRV_ORF
FT                   /note= "has similarity to KSHV interleukin (IL)-6 gene"
FT   CDS             complement (20777..21778)
FT                   /*tag= k
FT                   /product= "thymidylate synthase"
FT                   /label= RRV_ORF70
FT                   /note= "has similarity to KSHV ORF70"
FT   CDS             complement (22245..22592)
FT                   /*tag= l
FT                   /product= "RRV R3"
FT                   /note= "has similarity to KSHV K4 viral MIP gene"
FT   CDS             26846..27409
FT                   /*tag= m
FT                   /product= "Bcl2-homologue"
FT                   /label= RRV_ORF16
FT                   /note= "has similarity to KSHV ORF16"
FT   CDS             complememt (27515..29125)
FT                   /*tag= n
FT                   /label= RRV_ORF17
FT                   /note= "has similarity to KSHV ORF17"
FT   CDS             28998..29897
FT                   /*tag= o
FT                   /label= RRV_ORF18
FT                   /note= "has similarity to KSHV ORF18"
FT   CDS             complement (29905..31548)
FT                   /*tag= p
FT                   /product= "tegument protein
FT                   /label= RRV_ORF19
FT                   /note= "has similarity to KSHV ORF19"
FT   CDS             complement (31043..32095)
FT                   /*tag= q
FT                   /label= RRV_ORF20
FT                   /note= "has similarity to KSHV ORF20"
FT   CDS             32094..33767
FT                   /*tag= r
```

```
FT                   /product= "thymidine kinase"
FT                   /label= RRV_ORF21
FT                   /note= "has similarity to KSHV ORF21"
FT   CDS             33754..35868
FT                   /*tag=  s
FT                   /product= "glycoprotein H"
FT                   /label= RRV_ORF22
FT                   /note= "has similarity to KSHV ORF22"
FT   CDS             complement (35865..37073)
FT                   /*tag=  t
FT                   /label= RRV_ORF23
FT                   /note= "has similarity to KSHV ORF23"
FT   CDS             complement (37123..39321)
FT                   /*tag=  u
FT                   /label= RRV_ORF24
FT                   /note= "has similarity to KSHV ORF24"
FT   CDS             39323..43459
FT                   /*tag=  v
FT                   /product= "major capsid protein"
FT                   /label= RRV_ORF25
FT                   /note= "has similarity to KSHV ORF25"
FT   CDS             43491..44408
FT                   /*tag=  w
FT                   /product= "capsid protein"
FT                   /label= RRV_ORF26
FT                   /note= "has similarity to KSHV ORF26"
FT   CDS             44433..45242
FT                   /*tag=  x
FT                   /label= RRV_ORF27
FT                   /note= "has similarity to KSHV ORF27"
FT   CDS             45408..45683
FT                   /*tag=  y
FT                   /label= RRV_ORF28
FT                   /note= "has similarity to KSHV ORF28"
FT   CDS             complement (45733..46779)
FT                   /*tag=  z
FT                   /label= RRV_ORF29b
FT                   /note= "has similarity to KSHV ORF29b"
FT   CDS             46905..47135
FT                   /*tag=  aa
FT                   /label= RRV_ORF30
FT                   /note= "has similarity to KSHV ORF30"
FT   CDS             47093..47746
FT                   /*tag=  ab
FT                   /label= RRV_ORF31
FT                   /note= "has similarity to KSHV ORF31"
FT   CDS             47683..49077
FT                   /*tag=  ac
FT                   /label= RRV_ORF32
FT                   /note= "has similarity to KSHV ORF32"
FT   CDS             49049..50059
FT                   /*tag=  ad
FT                   /label= RRV_ORF33
FT                   /note= "has similarity to KSHV ORF33"
FT   CDS             complement (49977..50960)
FT                   /*tag=  ae
FT                   /label= RRV_ORF29a
FT                   /note= "has similarity to KSHV ORF29a"
FT   CDS             50959..51942
FT                   /*tag=  af
FT                   /label= RRV_ORF34
FT                   /note= "has similarity to KSHV ORF34"
FT   CDS             51923..52372
FT                   /*tag=  ag
FT                   /label= RRV_ORF35
FT                   /note= "has similarity to KSHV ORF35"
FT   CDS             52278..53585
FT                   /*tag=  ah
FT                   /product= "kinase"
FT                   /label= RRV_ORF36
FT                   /note= "has similarity to KSHV ORF36"
FT   CDS             53566..55008
FT                   /*tag=  ai
```

```
FT                   /product= "alkaline exonuclease"
FT                   /label= RRV_ORF37
FT                   /note= "has similarity to KSHV ORF37"
FT   CDS             54963..55172
FT                   /*tag=  aj
FT                   /label= RRV_ORF38
FT                   /note= "has similarity to KSHV ORF38"
FT   CDS             complement (55255..56391)
FT                   /*tag=  ak
FT                   /product= "glycoprotein M"
FT                   /label= RRV_ORF39
FT                   /note= "has similarity to KSHV ORF39"
FT   CDS             56526..57932
FT                   /*tag=  al
FT                   /product= "helicase/primase"
FT                   /label= RRV_ORF40
FT                   /note= "has similarity to KSHV ORF40"
FT   CDS             57917..58528
FT                   /*tag=  am
FT                   /product= "helicase/primase"
FT                   /label= RRV_ORF41
FT                   /note= "has similarity to KSHV ORF41"
FT   CDS             complement (58525..59343)
FT                   /*tag=  an
FT                   /label= RRV_ORF42
FT                   /note= "has similarity to KSHV ORF42"
FT   CDS             complement (59297..61027)
FT                   /*tag=  ao
FT                   /product= "capsid protein"
FT                   /label= RRV_ORF43
FT                   /note= "has similarity to KSHV ORF43"
```

```
  Query Match             12.5%;  Score 124.8;  DB 24;  Length 128139;
  Best Local Similarity   50.3%;  Pred. No. 8.4e-15;
  Matches  306;  Conservative    0;  Mismatches  302;  Indels    0;  Gaps    0;

Qy        258 AGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGG 317
              || ||||  ||   | | || ||  ||  ||       | ||| ||||   | | ||
Db     128003 AGCAGGCAAGCGGCCCCACTAGCAAGCAAGCAAGCAAGCCAGCAGGCAGCCCTACTAGCA 127944

Qy        318 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCT 377
              |  ||| || ||| || ||   |  ||   | | ||    |  ||| ||  || || ||
Db     127943 GCCCAGCAGCCCTACTAGCAGCCCAGCAGCCCTACTAGCAGCCCAGCAGCCCTACTAGCA 127884

Qy        378 TGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 437
              | || ||  || || || | || ||   | | || || | || |    | | || ||
Db     127883 GCCCAGGAGCCCTACTAGCAGCCCAGCAGCCCTACTAGCAGCCCAGGAGCCCTACTAGCA 127824

Qy        438 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 497
              | || ||   | | || || | || ||   | | || || | || ||   | | || ||
Db     127823 GCCCAGCAGCCCTACTAGCAGCCCAGCAGCCCTACTAGCAGCCCAGCAGCCCTACTAGCA 127764

Qy        498 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGAC 557
              | || ||   | | ||  |  |||  |  ||  ||   | | ||   | ||      |
Db     127763 GCCCAGCAGCCCTACTATCAAGCCATCAGGCAGGCGTACCGTCCCCTTAGCAAGCAAGCA 127704

Qy        558 AGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGC 617
                    |     || |   || | |   ||  ||  |  |||  |  || | |
Db     127703 GCCCTACTGTCCCCTCAGCAAGCAGGCCAGCGAGCAAGCAAGCAGGCCAGCGAGCAAGCA 127644

Qy        618 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 677
              || || | || || | |   || || | || || | |   || || | || || | |
Db     127643 AGCAGGCCAGCGAGCAAGCAAGCAGGCCAGCGAGCAAGCAAGCAGGCCAGCGAGCAAGCA 127584

Qy        678 TGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 737
              || || | || || | |    | || | || || | |   || || | || || ||  |
Db     127583 AGCAGGCCAGCGAGCAAGCAAGCAGGCCAGCGAGCAAGCAAGCAGGCCAGCGAGCAGCCC 127524

Qy        738 TCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 797
              | || |   || || ||  || |||  |    || | |   ||  |   || || ||| |
Db     127523 TACTAGCAAGCCAGCAGCCCTACTGTCCCCTCAGCAAGCAGGCCAGCGAGCAAGCAGGCC 127464

Qy        798 TGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 857
```

```
Db     127463 TACTGTCCCCTCAGCAAGCAGGCCTACTGTCCCCTCAGCAAGCAGGCCAGCGAGCAAGCA 127404

Qy        858 TGCCTGCT 865
              |||| ||
Db     127403 GGCCTACT 127396
```

RESULT 11
AAQ95095/c
ID AAQ95095 standard; DNA; 385 BP.
XX
AC AAQ95095;
XX
DT 02-OCT-1995 (first entry)
XX
DE Configuration detecting nucleic acid method target sequence.
XX
KW Probe; PCR; terminus; amplify; primer; configuration; ss.
XX
OS Synthetic.
XX
PN JP06343498-A.
XX
PD 20-DEC-1994.
XX
PF 03-JUN-1993; 93JP-0133640.
XX
PR 03-JUN-1993; 93JP-0133640.
XX
PA (CANO) CANON KK.
XX
DR WPI; 1995-069322/10.
XX
PT Nucleic acid probe - and method for detecting nucleic acid
XX
PS Example 1; Column 5; 6pp; Japanese.
XX
CC The sequence of the PCR amplified target region detected by probes
CC (AAQ95091-4). The probes used in a method for the detection of a
CC nucleic acid target sequence which has PCR-controlled termini and can be
CC amplified by PCR. The probes can be used to detect configuration.
XX
SQ Sequence 385 BP; 93 A; 97 C; 98 G; 97 T; 0 other;

```
  Query Match             9.8%;  Score 98.4;  DB 16;  Length 385;
  Best Local Similarity   55.6%;  Pred. No. 6.6e-10;
  Matches  189;  Conservative    0;  Mismatches  151;  Indels    0;  Gaps    0;

Qy        207 GACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 266
              |||    | |||  ||| ||| |||  ||  || ||| |||      |   | | || |
Db        381 GACGAGCGTGACGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCAT 322

Qy        267 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 326
              ||  || |||  || ||| |||      |   | | || ||| |||     | |  | |
Db        321 GCGCGCGCGCGCGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCAT 262

Qy        327 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 386
              || ||| |||  ||  || ||| |||      |   | | |  |||    |   ||
Db        261 GCATGCATGCATGCATGCATGCATGCATGCATGCATGCATTATATATATATA 202

Qy        387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 446
              |  |||  ||  || ||| ||| |||  ||  || ||| ||| |||  ||  || ||| |
Db        201 GCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCAT 142

Qy        447 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 506
              || |||  ||  || ||| ||| |||  ||  || |   | ||| |||    || ||| |
Db        141 GCATGCATGCATGCATGCATGCATGCATAACTACTTACTCTAGCGCATGCAT 82

Qy        507 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 546
              || |||  ||  || ||| ||| |||  ||  || ||| |
Db         81 GCATGCATGCATGCATGCATGCATGCATGCAT 42

RESULT 12
AAQ84040/c
ID AAQ84040 standard; DNA; 344 BP.
XX
AC AAQ84040;
XX
DT 03-OCT-1995 (first entry)
XX
DE Plasmid pCU19.
XX
KW Plasmid pUC19; DNA detection metod; ss.
XX
OS Synthetic.
XX
PN JP06343499-A.
XX
PD 20-DEC-1994.
XX
PF 04-JUN-1993; 93JP-0134615.
XX
PR 04-JUN-1993; 93JP-0134615.
XX
PA (CANO ) CANON KK.
XX
DR WPI; 1995-069323/10.
XX
PT Method for detecting nucleic acid - using nucleic acid probe
PT and PCR amplification
XX
PS Example 1; Page 4; 9pp; Japanese.
XX
CC AAQ84035 is the plasmid pUC19, used to demonstrate a new method for
CC the detection of DNA.
XX
SQ Sequence 344 BP; 84 A; 85 C; 86 G; 89 T; 0 other;

```
  Query Match   9.7%; Score 96.8; DB 16; Length 344;
  Best Local Similarity   55.3%; Pred. No. 1.3e-09;
  Matches 188; Conservative   0; Mismatches 152; Indels   0; Gaps   0;

Qy   207 GACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 266
              ||| | ||| ||| ||| ||| || || ||| ||| | | | || |
Db   340 GACGAGCGTGACGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCAT 281

Qy   267 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 326
              || || ||| || ||| ||| | | | || ||| ||| | | |
Db   280 GCGCGCGCGCGCGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCAT 221

Qy   327 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 386
              || ||| ||| || || ||| ||| | | | | ||| | ||
Db   220 GCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATTATATATATATA 161

Qy   387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 446
              | ||| || || ||| ||| ||| | || ||| ||| ||| || || ||| |
Db   160 GCATGCATGCATGCATGCATGCATGCATGGATGCATGCATGCATGCATGCATGCAT 101

Qy   447 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 506
              || ||| || || ||| ||| ||| || || | | ||| ||| || ||| |
Db   100 GCATGCATGCATGCATGCATGCATGCATGCATAACTACTTACTCTAGCGCATGCAT 41

Qy   507 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 546
              || ||| || || ||| ||| ||| || || ||| |
Db   40 GCATGCATGCATGCATGCATGCATGCATGCAT 1

RESULT 13
AAQ95095
ID AAQ95095 standard; DNA; 385 BP.
XX
AC AAQ95095;
XX
DT 02-OCT-1995 (first entry)

```
XX
DE   Configuration detecting nucleic acid method target sequence.
XX
KW   Probe; PCR; terminus; amplify; primer; configuration; ss.
XX
OS   Synthetic.
XX
PN   JP06343498-A.
XX
PD   20-DEC-1994.
XX
PF   03-JUN-1993;   93JP-0133640.
XX
PR   03-JUN-1993;   93JP-0133640.
XX
PA   (CANO ) CANON KK.
XX
DR   WPI; 1995-069322/10.
XX
PT   Nucleic acid probe - and method for detecting nucleic acid
XX
PS   Example 1; Column 5; 6pp; Japanese.
XX
CC   The sequence of the PCR amplified target region detected by probes
CC   (AAQ95091-4). The probes used in a method for the detection of a
CC   nucleic acid target sequence which has PCR-controlled termini and can be
CC   amplified by PCR. The probes can be used to detect configuration.
XX
SQ   Sequence 385 BP; 93 A; 97 C; 98 G; 97 T; 0 other;

  Query Match             9.7%;  Score 96.8;  DB 16;  Length 385;
  Best Local Similarity   56.6%;  Pred. No. 1.3e-09;
  Matches  179;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy        215 GTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 274
              | | |  ||| ||| |||  ||  || ||| |||    | |  |||||   ||| |||
Db         56 GCATGCATGCATGCATGCATGCATGCATGCGCTAGAGTAAGTAGTTATGCATGCAT 115

Qy        275 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 334
              ||  || ||| |||       |   | | || ||| |||      |   | | || ||| |
Db        116 GCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCAT 175

Qy        335 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTC 394
              ||  ||  || ||| |||      |  |             || |||  | | |  |||
Db.       176 GCATGCATGCATGCATGCATGCTATATATATATAATGCATGCATGCATGCATGCAT 235

Qy        395 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 454
              ||  || ||| ||| |||  ||  || ||| ||| |||  ||  || ||| ||| |||
Db        236 GCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCAT 295

Qy        455 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 514
              ||  || ||| |||  || |||  || ||| ||| |||  ||  || ||| ||| |||
Db        296 GCATGCATGCATGCGCGCGCGCGCGCATGCATGCATGCATGCATGCATGCATGCAT 355

Qy        515 GCTGGCCTGCTTGCTT 530
              ||  || ||| ||| |
Db        356 GCATGCATGCATGCGT 371


RESULT 14
AAK79692
ID   AAK79692 standard; DNA; 14212 BP.
XX
AC   AAK79692;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34504.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
```

PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 34504; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 14212 BP; 3809 A; 3421 C; 3440 G; 3542 T; 0 other;

Query Match 9.6%; Score 96.4; DB 22; Length 14212;
Best Local Similarity 48.7%; Pred. No. 1.7e-09;
Matches 294; Conservative 0; Mismatches 306; Indels 4; Gaps 1;

```
Qy       283 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 342
            ||||  ||     |  |||  || |  ||      ||||||   | ||| ||  | | |
Db      6579 GCTTTGTGGACAAAATGTACACCAGGCTGAGATATGGACAGGTTGACTGGTTAGTAGGTT 6638

Qy       343 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCT 402
            | |  ||| |  |  |  ||   || |  | | |||   |   || || | | | |  |
Db      6639 GGTTAGTTGGTTGGCTGGTAGGTTGGTTAGTTGGTTGACTGGTAGGTTGATAGGTAGGTT 6698

Qy       403 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 462
            |  |  ||| | | |    || ||  | | | |   |  |  ||| ||  | | | | |
Db      6699 GGCTAGTTGATTGGTAAGTTGGTTCGTAGGTTGGGAGGTTAGTTGGTTAGTTGGTTGGTT 6758

Qy       463 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 522
            |||| ||| |  | | |  |  |  ||| | | |    || ||| ||| | | | |
Db      6759 GCTTGGTTGGCAGGTTGGTTTGTATGTTGGTAGGTTAGTTGATTGGTTGGTAGGTTGGTA 6818

Qy       523 GCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTG----CTTGCTA 578
            | |||  | | | ||| |  ||| | ||  |   |   || || |||     ||| ||
Db      6819 GATTGGCAGGTTGGTGGGTTAGTTGGTTAGTTGGTTGGTAGGTTGGTTGGCAAGTTGATA 6878

Qy       579 AGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 638
            |  |   |  | || ||||  |||||||  || || |  | || |||   || |
Db      6879 GGTTGGTTGGTGGGTTGGTTGATAAGTTGATAGGTTGGTTGGTAAGTTGATAGGTTGGTT 6938

Qy       639 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 698
            | |||| || | |    | || | || || | |   || || | |   |   |  | |
Db      6939 GGTGGGTTGGTTGATAAGTTGATAGGTTGGTTGGTAGGTTGGTTGTTAGGGTGGTAGATT 6998

Qy       699 CCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCT 758
               || || | |   |  |  | ||  | | |   || |    || || | |     |
Db      6999 GGCAGGTTGGTTGGTAGTTTTGTTGGTAGTTTGGTTGGTTAGCTGGTTGGTTGGTTAGTG 7058

Qy       759 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 818
            | || || |||   |  || | |  || | |   || |  |  | || |||    |
Db      7059 AGTTGGTTGGTAGATAGATTGGTAGATTGGTTGGTTGGTTAGTTTGTTGGTAGGTTAGTA 7118

Qy       819 GCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTG 878
            |  |  |   ||| ||  ||| ||| |||   ||  | ||| |||  |  |   || |
Db      7119 GGTTTGTTAGTTGGTTAGTTGATTGGTAGGTTGGTTGGTTGACTGAAAGGACCCCTTCCT 7178
```

```
Qy        879 TGCA 882
              | ||
Db       7179 TACA 7182


RESULT 15
AAK80024
ID   AAK80024 standard; DNA; 14212 BP.
XX
AC   AAK80024;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34836.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
```

```
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 34836; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 14212 BP; 3809 A; 3421 C; 3440 G; 3542 T; 0 other;

  Query Match             9.6%;  Score 96.4;  DB 22;  Length 14212;
  Best Local Similarity   48.7%;  Pred. No. 1.7e-09;
  Matches  294;  Conservative    0;  Mismatches  306;  Indels    4;  Gaps    1;

Qy        283 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 342
Db       6579 GCTTTGTGGACAAAATGTACACCAGGCTGAGATATGGACAGGTTGACTGGTTAGTAGGTT 6638

Qy        343 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCT 402
Db       6639 GGTTAGTTGGTTGGCTGGTAGGTTGGTTAGTTGGTTGACTGGTAGGTTGATAGGTAGGTT 6698

Qy        403 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 462
Db       6699 GGCTAGTTGATTGGTAAGTTGGTTCGTAGGTTGGGAGGTTAGTTGGTTAGTTGGTTGGTT 6758

Qy        463 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 522
Db       6759 GCTTGGTTGGCAGGTTGGTTTGTATGTTGGTAGGTTAGTTGATTGGTTGGTAGGTTGGTA 6818

Qy        523 GCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTG----CTTGCTA 578
Db       6819 GATTGGCAGGTTGGTGGGTTAGTTGGTTAGTTGGTTGGTAGGTTGGTTGGCAAGTTGATA 6878

Qy        579 AGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 638
Db       6879 GGTTGGTTGGTGGGTTGGTTGATAAGTTGATAGGTTGGTTGGTAAGTTGATAGGTTGGTT 6938

Qy        639 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 698
Db       6939 GGTGGGTTGGTTGATAAGTTGATAGGTTGGTTGGTAGGTTGGTTGTTAGGGTGGTAGATT 6998

Qy        699 CCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCT 758
Db       6999 GGCAGGTTGGTTGGTAGTTTGTTGGTAGTTTGGTTGGTTAGCTGGTTGGTTGGTTAGTG 7058

Qy        759 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 818
Db       7059 AGTTGGTTGGTAGATAGATTGGTAGATTGGTTGGTTGGTTAGTTTGTTGGTAGGTTAGTA 7118

Qy        819 GCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTG 878
Db       7119 GGTTTGTTAGTTGGTTAGTTGATTGGTAGGTTGGTTGGTTGACTGAAAGGACCCCTTCCT 7178

Qy        879 TGCA 882
Db       7179 TACA 7182


RESULT 16
AAK85921
ID   AAK85921 standard; DNA; 14212 BP.
XX
AC   AAK85921;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40733.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
```

PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 40733; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 14212 BP; 3809 A; 3421 C; 3440 G; 3542 T; 0 other;

Query Match 9.6%; Score 96.4; DB 22; Length 14212;
Best Local Similarity 48.7%; Pred. No. 1.7e-09;
Matches 294; Conservative 0; Mismatches 306; Indels 4; Gaps 1;

```
Qy        283 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 342
Db       6579 GCTTTGTGGACAAAATGTACACCAGGCTGAGATATGGACAGGTTGACTGGTTAGTAGGTT 6638

Qy        343 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCT 402
Db       6639 GGTTAGTTGGTTGGCTGGTAGGTTGGTTAGTTGGTTGACTGGTAGGTTGATAGGTAGGTT 6698

Qy        403 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 462
Db       6699 GGCTAGTTGATTGGTAAGTTGGTTCGTAGGTTGGGAGGTTAGTTGGTTAGTTGGTTGGTT 6758

Qy        463 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 522
Db       6759 GCTTGGTTGGCAGGTTGGTTTGTATGTTGGTAGGTTAGTTGATTGGTTGGTAGGTTGGTA 6818

Qy        523 GCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTG----CTTGCTA 578
Db       6819 GATTGGCAGGTTGGTGGGTTAGTTGGTTAGTTGGTTGGTAGGTTGGTTGGCAAGTTGATA 6878

Qy        579 AGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 638
Db       6879 GGTTGGTTGGTGGGTTGGTTGATAAGTTGATAGGTTGGTTGGTAAGTTGATAGGTTGGTT 6938

Qy        639 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 698
Db       6939 GGTGGGTTGGTTGATAAGTTGATAGGTTGGTTGGTAGGTTGGTTGTTAGGGTGGTAGATT 6998

Qy        699 CCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCT 758
Db       6999 GGCAGGTTGGTTGGTAGTTTTGTTGGTAGTTTGGTTGGTTAGCTGGTTGGTTGGTTAGTG 7058

Qy        759 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 818
Db       7059 AGTTGGTTGGTAGATAGATTGGTAGATTGGTTGGTTGGTTAGTTTGTTGGTAGGTTAGTA 7118

Qy        819 GCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTG 878
Db       7119 GGTTTGTTAGTTGGTTAGTTGATTGGTAGGTTGGTTGGTTGACTGAAAGGACCCCTTCCT 7178

Qy        879 TGCA 882
Db       7179 TACA 7182
```

RESULT 17
AAQ84040
ID AAQ84040 standard; DNA; 344 BP.
XX
AC AAQ84040;
XX
DT 03-OCT-1995 (first entry)
XX
DE Plasmid pCU19.
XX
KW Plasmid pUC19; DNA detection metod; ss.
XX
OS Synthetic.
XX
PN JP06343499-A.
XX
PD 20-DEC-1994.
XX
PF 04-JUN-1993; 93JP-0134615.
XX
PR 04-JUN-1993; 93JP-0134615.
XX
PA (CANO) CANON KK.
XX
DR WPI; 1995-069323/10.
XX
PT Method for detecting nucleic acid - using nucleic acid probe
PT and PCR amplification
XX
PS Example 1; Page 4; 9pp; Japanese.
XX
CC AAQ84035 is the plasmid pUC19, used to demonstrate a new method for
CC the detection of DNA.
XX
SQ Sequence 344 BP; 84 A; 85 C; 86 G; 89 T; 0 other;

Query Match 9.5%; Score 95.2; DB 16; Length 344;
Best Local Similarity 56.3%; Pred. No. 2.6e-09;
Matches 178; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

```
Qy        215 GTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 274
Db         15 GCATGCATGCATGCATGCATGCATGCATGCATGCGCTAGAGTAAGTAGTTATGCATGCAT 74

Qy        275 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 334
Db         75 GCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATCCATCCAT 134

Qy        335 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTC 394
Db        135 GCATGCATGCATGCATGCATGCATGCTATATATATATAATGCATGCATGCATGCATGCAT 194

Qy        395 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 454
Db        195 GCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCATGCAT 254

Qy        455 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 514
Db        255 GCATGCATGCATGCGCGCGCGCGCGCATGCATGCATGCATGCATGCATGCATGCATGCAT 314

Qy        515 GCTGGCCTGCTTGCTT 530
Db        315 GCATGCATGCATGCGT 330
```

RESULT 18
ABQ31996
ID ABQ31996 standard; DNA; 1172 BP.
XX
AC ABQ31996;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 18587.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -

```
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 1172 BP; 0 A; 583 C; 586 G; 3 T; 0 other;

  Query Match             9.4%;  Score 93.8;  DB 24;  Length 1172;
  Best Local Similarity   47.3%;  Pred. No. 5e-09;
  Matches  284;  Conservative    0;  Mismatches  317;  Indels    0;  Gaps    0;

Qy          4 CGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTC 63
              ||| ||     ||| |   ||| ||   | | | | | |||   | ||   |   |
Db        571 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 630

Qy         64 TCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGC 123
                |||||   || | | | |  || |  |  | ||||    ||| |  |  ||||| |
Db        631 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 690

Qy        124 GCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGG 183
                |   | | || | |  || | |  || |    |||| | |     | | |  |  |
Db        691 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 750

Qy        184 GGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTG 243
                |  ||| | ||| ||  | | |    || |  ||   ||  ||  || ||| ||  |
Db        751 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 810

Qy        244 CTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGG 303
                |  ||    | |   |   ||  ||  || ||| ||  ||  ||    | |   |   |
Db        811 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 870

Qy        304 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 363
                |  ||  ||    | |   |   ||  ||  || ||| ||  ||  ||    | |   |
Db        871 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 930

Qy        364 TAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 423
                |   ||  ||  ||  |   |   ||  ||| ||  ||  || |||  ||  ||  |
Db        931 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGCGCGCGCGCGCGCGCG 990

Qy        424 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 483
                |  ||  ||  |||  ||  ||  ||  ||  |||  ||  ||  ||  ||  |||  ||  |
Db        991 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1050

Qy        484 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 543
                |  ||  ||  |||  ||  ||  ||  ||  |||  ||  ||  ||  |||  ||  |
Db       1051 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGCGCGCGCGCGCGCGCGCGCG 1110

Qy        544 CTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGG 603
                |  ||    | |   ||  |   ||  ||  ||    | |   | |||  ||  |   |
Db       1111 CGCGCGCGCGCGCGCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1170

Qy        604 C 604
              |
Db       1171 C 1171
```

```
RESULT 19
ABQ31997/c
ID   ABQ31997 standard; DNA; 1172 BP.
XX
AC   ABQ31997;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 18588.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 1172 BP; 3 A; 586 C; 583 G; 0 U; 0 other;

  Query Match             9.4%;  Score 93.8;  DB 24;  Length 1172;
  Best Local Similarity   47.3%;  Pred. No. 5e-09;
  Matches  284;  Conservative    0;  Mismatches  317;  Indels    0;  Gaps    0;

Qy          4 CGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTC 63
              ||| ||     ||| |   ||| ||   | | | | | |||   | ||   |   |
Db        602 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 543

Qy         64 TCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGC 123
                |||||   || | | | |  || |  |  | ||||    ||| |  |  ||||| |
```

```
Db        542 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 483

Qy        124 GCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGG 183

Db        482 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 423

Qy        184 GGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTG 243

Db        422 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 363

Qy        244 CTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGG 303

Db        362 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 303

Qy        304 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 363

Db        302 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 243

Qy        364 TAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 423

Db        242 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGCGCGCGCGCGCGCGCG 183

Qy        424 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 483

Db        182 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 123

Qy        484 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 543

Db        122 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGCGCGCGCGCGCGCGCGCGCGCGCG 63

Qy        544 CTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGG 603

Db         62 CGCGCGCGCGCGCGCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3

Qy        604 C 604

Db          2 C 2
```

RESULT 20
ABQ31994/c
ID ABQ31994 standard; DNA; 1172 BP.
XX
AC ABQ31994;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 18585.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 1172 BP; 3 A; 583 C; 583 G; 3 T; 0 other;

```
  Query Match             9.2%;  Score 92.2;  DB 24;  Length 1172;
  Best Local Similarity   47.1%;  Pred. No. 9.9e-09;
  Matches  283;  Conservative    0;  Mismatches  318;  Indels    0;  Gaps    0;

Qy          4 CGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTC 63

Db        784 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 725

Qy         64 TCCGCGCCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGC 123

Db        724 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 665

Qy        124 GCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGG 183

Db        664 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 605

Qy        184 GGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTG 243

Db        604 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 545

Qy        244 CTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGG 303

Db        544 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 485

Qy        304 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 363

Db        484 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 425

Qy        364 TAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 423

Db        424 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 365

Qy        424 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 483

Db        364 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 305

Qy        484 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 543

Db        304 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 245

Qy        544 CTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGG 603

Db        244 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCG 185

Qy        604 C 604
```

```
Db        184 C 184


RESULT 21
ABQ31995
ID   ABQ31995 standard; DNA; 1172 BP.
XX
AC   ABQ31995;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 18586.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 1172 BP; 3 A; 583 C; 583 G; 3 T; 0 other;

  Query Match             9.2%;  Score 92.2;  DB 24;  Length 1172;
  Best Local Similarity   47.1%;  Pred. No. 9.9e-09;
  Matches  283;  Conservative    0;  Mismatches  318;  Indels    0;  Gaps    0;

Qy          4 CGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTC 63
              ||| ||     ||| |   |||  ||   | | | | | |||   | ||   |   |
Db        389 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 448

Qy         64 TCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGC 123
                ||||   || | | | |  || |  |  | ||||    ||| |  |  ||||| |
Db        449 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 508

Qy        124 GCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGG 183
                |   | | || | |  || | |  || |    |||| | |     | |  | |
Db        509 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 568

Qy        184 GGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTG 243
               |  ||| | ||| ||  | | |    || |  ||   ||  ||  || ||| ||  |
Db        569 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 628

Qy        244 CTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGG 303
               |  ||    | |   |   ||  ||  || |||  ||  ||  ||    | |   |  |
Db        629 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 688

Qy        304 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 363
               |  ||  ||    | |   |   ||  ||  || |||  ||  ||  ||    | |   |
Db        689 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 748

Qy        364 TAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 423
               |   ||  ||  ||  |   |   || |||  ||  ||  ||  || |||  ||  |
Db        749 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 808

Qy        424 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 483
               |  ||  || |||  ||  ||  ||  || |||  ||  ||  ||  || |||  ||  |
Db        809 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 868

Qy        484 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 543
               |  ||  || |||  ||  ||  ||  || |||  ||  ||  ||  || |||  ||  |
Db        869 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 928

Qy        544 CTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGG 603
               |  ||    | |   |   |   ||  ||  ||    | |   ||||||  ||  |   |
Db        929 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCG 988

Qy        604 C 604
               |
Db        989 C 989


RESULT 22
ABZ39851/c
ID   ABZ39851 standard; DNA; 2130 BP.
XX
AC   ABZ39851;
XX
DT   07-MAR-2003  (first entry)
XX
DE   N. gonorrhoeae nucleotide sequence SEQ ID 4291.
XX
KW   Antibacterial; infection; vaccine; gene therapy; gene; ds.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO200279243-A2.
XX
PD   10-OCT-2002.
XX
PF   12-FEB-2002; 2002WO-IB02069.
XX
PR   12-FEB-2001; 2001GB-0003424.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Fontana MR,  Pizza M,  Masignani V,  Monaci E;
XX
DR   WPI; 2003-058415/05.
DR   P-PSDB; ABP78881.
XX
PT   New protein from Neisseria gonorrheae, useful for the manufacture of a
PT   medicament for treating or preventing N. gonorrheae infection  -
XX
PS   Disclosure; Page 502; 815pp; English.
```

```
XX
CC   The present invention relates to proteins from Neisseria gonorrhoeae.
CC   Also disclosed are the nucleic acid molecules encoding the proteins and
CC   antibodies that specifically bind to the proteins. The composition
CC   comprising the protein, nucleic acid or antibody is useful for the
CC   manufacture of a medicament for treating or preventing N. gonorrhoeae
CC   infection, this may be in the form of a vaccine or gene therapy.
CC   Sequences given in records ABZ37706-ABZ42016 represent nucleic acid
CC   molecules of the invention.
XX
SQ   Sequence 2130 BP; 792 A; 394 C; 401 G; 543 T; 0 other;

  Query Match             9.2%;  Score 91.6;  DB 25;  Length 2130;
  Best Local Similarity   71.2%;  Pred. No. 1.3e-08;
  Matches  121;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        376 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 435
              || | || |  ||    |  ||||  ||| ||| ||| ||||||  ||| ||| ||| |
Db        443 CTGGTTAATTTGGTGTATAATTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 384

Qy        436 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 495
              |||||  ||| ||| ||| ||||||  ||| ||| ||| ||||||  ||| ||| ||| |
Db        383 CTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 324

Qy        496 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 545
              |||||  ||| ||| ||| ||||||  ||| ||| ||| ||||||  |||
Db        323 CTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 274


RESULT 23
ABQ31996/c
ID   ABQ31996 standard; DNA; 1172 BP.
XX
AC   ABQ31996;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 18587.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 1172 BP; 0 A; 583 C; 586 G; 3 T; 0 other;

  Query Match             9.1%;  Score 90.8;  DB 24;  Length 1172;
  Best Local Similarity   47.1%;  Pred. No. 1.8e-08;
  Matches  278;  Conservative    0;  Mismatches  312;  Indels    0;  Gaps    0;

Qy         15 GCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCTCCGCGCCGCC 74
              |||  |   |||  ||   | | | | |||||   | ||   |   |    |||||   |
Db       1117 GCGCGCGCGCGCGCGCGCGCGCGCGCGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1058

Qy         75 GGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGACCGAGGG 134
              | | | | |  || |  |  | ||||    |||  |  | |||||  |       |    |
Db       1057 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 998

Qy        135 CCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGTGT 194
               | || | |  || | |  || | |  |||| | |      | |  |  | |   ||| |
Db        997 GCGCGCGCGCGCGCGCGCGCGCGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 938

Qy        195 GCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGG 254
              |||  ||  |  | |    || |  ||   ||  ||  ||  ||  |||  ||  ||  |
Db        937 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 878

Qy        255 GACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTG 314
              |   |   ||  ||  || |||  ||  ||  ||    | |   |   ||  ||  ||
Db        877 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 818

Qy        315 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTG 374
               | |   |  ||  ||  || |||  ||  ||  ||    | |   |   |   ||
Db        817 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 758

Qy        375 GCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 434
              ||  ||  |   |   || |||  ||  ||  ||  || |||  ||  ||  ||  || |
Db        757 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 698

Qy        435 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 494
              ||  ||  ||  ||  || |||  ||  ||  ||  || |||  ||  ||  ||  || |
Db        697 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 638

Qy        495 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 554
              ||  ||  ||  ||  || |||  ||  ||  ||  || |||  ||  ||  ||    |
Db        637 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 578

Qy        555 GACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGC 604
              |   |   |   ||  ||  ||    | |   |  |||  ||  |   ||
Db        577 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 528


RESULT 24
ABQ31997
ID   ABQ31997 standard; DNA; 1172 BP.
XX
AC   ABQ31997;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 18588.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
```

```
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 1172 BP; 3 A; 586 C; 583 G; 0 U; 0 other;

  Query Match             9.1%;  Score 90.8;  DB 24;  Length 1172;
  Best Local Similarity   47.1%;  Pred. No. 1.8e-08;
  Matches  278;  Conservative    0;  Mismatches  312;  Indels    0;  Gaps    0;

Qy         15 GCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCTCCGCGCCGCC 74
              |||  |   |||  ||   | | | | |||||   | ||   |   |    |||||   |
Db         56 GCGCGCGCGCGCGCGCGCGCGCGCGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 115

Qy         75 GGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGACCGAGGG 134
              | | | | |  || | |  | ||||    ||| |  |  ||||| |      |    |
Db        116 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 175

Qy        135 CCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGTGT 194
              | || | |  || | |  || | |  |||| | |     | |  |  | |   ||| |
Db        176 GCGCGCGCGCGCGCGCGCGCGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 235

Qy        195 GCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGG 254
              |||  ||  | | |    || |  ||   ||  ||  || |||  ||  ||  ||    |
Db        236 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 295

Qy        255 GACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTG 314
              |   |   ||  ||  || |||  ||  ||  ||    | |   |   ||  ||  ||
Db        296 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 355

Qy        315 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTG 374
              | |   |   ||  ||  || |||  ||  ||  ||    | |   |   |   ||
Db        356 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 415

Qy        375 GCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 434
              ||  ||  |   |   || |||  ||  ||  ||  || |||  ||  ||  ||  || |
Db        416 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 475

Qy        435 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 494
              ||  ||  ||  ||  || |||  ||  ||  ||  || |||  ||  ||  ||  || |
Db        476 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 535

Qy        495 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 554
              ||  ||  ||  ||  || |||  ||  ||  ||  || |||  ||  ||  ||    |
Db        536 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 595

Qy        555 GACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGC 604
              |   |   |  ||  ||  ||    | |   |  |||  ||  |   ||
Db        596 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 645


RESULT 25
ABQ31994
ID   ABQ31994 standard; DNA; 1172 BP.
XX
AC   ABQ31994;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 18585.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
```

```
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 1172 BP; 3 A; 583 C; 583 G; 3 T; 0 other;

  Query Match             9.1%;  Score 90.6;  DB 24;  Length 1172;
  Best Local Similarity   46.9%;  Pred. No. 2e-08;
  Matches  282;  Conservative    0;  Mismatches  319;  Indels    0;  Gaps    0;

Qy          4 CGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTC 63
              ||| ||     ||| |   ||| ||   | | | | | |||   | ||   |   |
Db         77 CGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 136

Qy         64 TCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGC 123
                |||||   || | | | |  || |  |  | ||||    ||| |  |  |||||  |
Db        137 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 196

Qy        124 GCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGG 183
                ||    |  | || | |  || | |  || |    |||| | |      | |  |  |
Db        197 CGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 256

Qy        184 GGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTG 243
               |   ||| | ||| ||  | | |    || |  ||   ||  ||  || |||  ||  |
Db        257 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 316

Qy        244 CTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGG 303
              |  ||    | |   |   ||  ||  || ||| ||  ||  ||    | |   |   |
Db        317 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 376

Qy        304 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 363
              |  ||  ||    | |   |   ||  ||  || ||| ||  ||  ||    | |   |
Db        377 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 436

Qy        364 TAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 423
                |  ||  ||  ||  |   |   || |||  ||  ||  ||  || |||  ||  |
Db        437 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 496

Qy        424 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 483
              |  ||  || |||  ||  ||  ||  || |||  ||  ||  ||  || |||  ||  |
Db        497 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 556

Qy        484 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 543
                |  ||  || |||  ||  ||  ||  || |||  ||  ||  ||  || |||  ||  |
Db        557 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 616

Qy        544 CTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGG 603
              |  ||    | |   |   |  ||  ||  ||    | |   |  |||  ||  |   |
Db        617 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 676

Qy        604 C 604
              |
Db        677 C 677


RESULT 26
ABQ31995/c
ID   ABQ31995 standard; DNA; 1172 BP.
XX
AC   ABQ31995;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 18586.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 1172 BP; 3 A; 583 C; 583 G; 3 T; 0 other;

  Query Match             9.1%;  Score 90.6;  DB 24;  Length 1172;
  Best Local Similarity   46.9%;  Pred. No. 2e-08;
  Matches  282;  Conservative    0;  Mismatches  319;  Indels    0;  Gaps    0;

Qy          4 CGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTC 63
              ||| ||     ||| |   ||| ||   | | | | | |||   | ||   |   |
Db       1096 CGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1037

Qy         64 TCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGC 123
                |||||   || | | | |  || |  |  | ||||    ||| |  |  |||||  |
Db       1036 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 977

Qy        124 GCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGG 183
                ||    |  | || | |  || | |  || |    |||| | |      | |  |  |
Db        976 CGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 917

Qy        184 GGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTG 243
               |   ||| | ||| ||  | | |    || |  ||   ||  ||  || |||  ||  |
Db        916 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 857

Qy        244 CTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGG 303
              |  ||    | |   |   ||  ||  || ||| ||  ||  ||    | |   |   |
Db        856 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 797

Qy        304 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 363
              |  ||  ||    | |   |   ||  ||  || ||| ||  ||  ||    | |   |
```

```
Db        796 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 737

Qy        364 TAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 423
              |  ||  ||  ||  |   |  || ||| ||  ||  ||  || |||  ||  |
Db        736 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 677

Qy        424 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 483
              |  ||  || |||  ||  ||  ||  || |||  ||  ||  ||  || |||  ||  |
Db        676 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 617

Qy        484 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 543
              |  ||  || |||  ||  ||  ||  || |||  ||  ||  ||  || |||  ||  |
Db        616 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 557

Qy        544 CTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGG 603
              |  ||    | |   |   |   ||  ||  ||    | |   |  ||| ||  |   |
Db        556 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 497

Qy        604 C 604
              |
Db        496 C 496
```

RESULT 27
ABA15389/c
ID ABA15389 standard; DNA; 2855 BP.
XX
AC ABA15389;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 7720.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.

```
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250391.
PR   01-DEC-2000; 2000US-0251160.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system
PT   cancers and metastases -
XX
PS   Disclosure; SEQ ID NO 7720; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2855 BP; 592 A; 829 C; 944 G; 490 T; 0 other;
```

```
  Query Match             8.7%;  Score 87.2;  DB 22;  Length 2855;
  Best Local Similarity   54.3%;  Pred. No. 8.9e-08;
  Matches  176;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy        225 CTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGC 284
              || |   ||| || | |  | |||   ||   | ||  |     ||||  |||   ||||
Db       1938 CTCCTGACTCCCTTGTCCACGTGCAAGGGCACCTGTGAGTGCATTTGCATGCTTATCTGC 1879

Qy        285 TTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGC 344
               |||| |   |   |   |  ||| | ||||   |   |   |  ||| |||||     |
Db       1878 ATGCTCATTTGTATGCTTGTTTGCCTACTGATCTGCATGCTCGTTTGCATGCTCATCTAC 1819

Qy        345 CTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGC 404
               |||| | |   | |    |  |  |||| | |||| ||     ||||  ||| |  |||
Db       1818 TTGCTCGTTTGCGTGGTCATCTGCGTGCTCGTTTGCGAGCTCATCTGCGTGCTCGTTTGC 1759

Qy        405 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 464
               |||| |   ||  ||  |  ||  |||||    || |||| |  ||   |||   ||||
Db       1758 GTGCTCGTCTGCGTGCTCGTCTGTGTGCTCATCTGCGTGCTCGTCTGTGTGCTCATCTGC 1699

Qy        465 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 524
               ||||     ||  || || |||| |||||    |  |||||| |||   ||| | ||||
Db       1698 ATGCTCATCTGCGTGCTTGTTTGCATGCTCATCTGTGTGCTTGTTTGTGTGCTCGTCTGC 1639

Qy        525 TTGCTTGCTCGCTGGCCTGCTTGC 548
               |||||| | ||  ||    ||||
Db       1638 GTGCTTGTTTGCATGCTCATTTGC 1615
```

```
RESULT 28
AAT42063
ID   AAT42063 standard; DNA; 1830121 BP.
XX
AC   AAT42063;
XX
DT   14-SEP-1999  (first entry)
XX
DE   Haemophilus influenzae complete genome sequence.
XX
KW   Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW   expression modulating fragment; regulation; gene expression; vector;
KW   organism; open reading frame; ORF; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO9633276-A1.
XX
PD   24-OCT-1996.
XX
PF   22-APR-1996;   96WO-US05320.
XX
PR   07-JUN-1995;   95US-0487429.
PR   21-APR-1995;   95US-0426787.
PR   07-JUN-1995;   95US-0476102.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (UYJO ) UNIV JOHNS HOPKINS.
XX
PI   Adams MD,  Fleischmann RD,  Smith HO,  Venter JC,  White O;
XX
DR   WPI; 1996-485782/48.
XX
PT   Haemophilus influenzae Rd genome recorded on computer readable
PT   medium  - useful for identifying commercially important nucleic acid
PT   fragments by homology searching
XX
PS   Claim 1; Page 77.2-77.1091; 1291pp; English.
XX
CC   This sequence represents the complete genome sequence of the bacterium
CC   Haemophilus influenzae strain Rd.  The invention relates to a computer
CC   readable medium (CRM) having recorded upon it the complete H.influenzae
CC   nucleotide sequence (I), a representative fragment of (I) or a nucleotide
CC   sequence at least 99% identical to (I).  By providing the full-length
```

```
CC   genomic sequence in a computer readable form, it is possible to identify
CC   commercially important nucleic acid fragments and expression modulating
CC   fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC   regulate the expression of a nucleic acid molecule.  Vectors and altered
CC   organisms comprising the predicted ORFS can be used to produce any of the
CC   polypeptide fragments of the H. influenzae Rd genome.
XX
SQ   Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

  Query Match             8.0%;  Score 79.8;  DB 17;  Length 1830121;
  Best Local Similarity   84.1%;  Pred. No. 2.6e-06;
  Matches   90;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy        405 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 464
              ||||||||| ||| || ||||||||||||| ||| || ||||||||||||| ||| || |||
Db    1543129 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 1543188

Qy        465 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 511
              ||||||||| ||| || |||||||||||| ||| || |||||  |||
Db    1543189 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTAATTG 1543235


RESULT 29
AAK89506/c
ID   AAK89506 standard; DNA; 15682 BP.
XX
AC   AAK89506;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human digestive system antigen genomic sequence SEQ ID NO: 3082.
XX
KW   Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW   ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW   digestive system disorder; Meckel's diverticulum; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155314-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01324.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
```

```
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-502630/55.
XX
PT   Polynucleotides encoding digestive system antigens, useful for
PT   diagnosing, treating, preventing and/or prognozing disorders of the
PT   digestive system, particularly cancer and cancer metastases -
XX
PS   Disclosure; SEQ ID NO 3082; 986pp; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human digestive system antigens. These can be used in the
CC   diagnosis, treatment and prevention of digestive system disorders,
CC   including cancer, Meckel's diverticulum, bacterial or parasitic
CC   infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC   ulcerative colitis. The present sequence is a genomic DNA fragment
CC   encoding a digestive system antigen of the invention.
XX
SQ   Sequence 15682 BP; 3797 A; 4055 C; 4103 G; 3727 T; 0 other;

  Query Match             7.9%;  Score 78.8;  DB 22;  Length 15682;
  Best Local Similarity   50.5%;  Pred. No. 3.5e-06;
  Matches  191;  Conservative    0;  Mismatches  187;  Indels    0;  Gaps    0;

Qy        221 CTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGC 280
              ||||||||||| |   |  |||||| | |      | | |   |||| | | |  || ||
Db      12169 CTGCCTGCCTTCCTTCCTGCCTGCCTTCCTTCCTGCCTGCCTGCCTTCCTTCCTGCCTGC 12110

Qy        281 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGC 340
              || | | |      | |      ||| | |||       |     |||||| |||   |
Db      12109 CTTCCTTCCTGCCTGCCTTCCTTCCTTCCTGCCTGCCTTCCTTCCTGCCTGCCTGCCTTC 12050

Qy        341 TGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGC 400
                |||| | |||             | ||||  || | |    |   ||||  ||  ||
Db      12049 CTGCCTTCCTGCCTTCATTCATTCCTGCCTGCCTGCCTTCATTCATTCCTGCCTGCCTGC 11990

Qy        401 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 460
              || | | | | |  ||  |||| | | | | |  ||  |||| | | | ||| |||   |
Db      11989 CTTCCTTCCTTCCTGCCCGCCTTCCTTCCTTCCTGCCCGCCTTCCTTCCTGCCCGCCTTC 11930

Qy        461 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 520
              || | | |||  ||   ||||| | | |||  ||  |||| | | | | |  ||  ||
Db      11929 CTTCCTTCCTGCCTGCCTTCCTGCCTTCCTGCCTGCCTGCCTTCATTCATTCCTGCCTGC 11870

Qy        521 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAG 580
              |||| | | | |   | ||||| | |       |       || | ||| |||
Db      11869 CTGCCTTCATTCATTCCTGCCTGCCTTCCTTCATTCCTTCCTTCCTTCCTGCCTGCCTGC 11810

Qy        581 GGGACGGTACGCCTGCCT 598
                | |   ||||||||
Db      11809 CTTCCTGCCTGCCTGCCT 11792


RESULT 30
AAL40273/c
ID   AAL40273 standard; DNA; 169 BP.
XX
AC   AAL40273;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 35.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   N-PSDB; AAO21920, AAO21921.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear  -
XX
PS   Example 1; Fig 3; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
```

```
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 169 BP; 69 A; 38 C; 33 G; 29 T; 0 other;

  Query Match             7.8%;  Score 78;  DB 24;  Length 169;
  Best Local Similarity   85.3%;  Pred. No. 4.4e-06;
  Matches   87;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy        389 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 448
              ||||| ||| || |||||||||||| ||| || |||||||||||| ||| || |||||||
Db        112 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 53

Qy        449 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              ||||| ||| || |||||||||||| ||| || |||||||||
Db         52 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 11


RESULT 31
ABA19579
ID   ABA19579 standard; DNA; 7202 BP.
XX
AC   ABA19579;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 11910.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   20-OCT-2000; 2000US-0242221.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
```

```
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250391.
PR   01-DEC-2000; 2000US-0251160.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system
PT   cancers and metastases -
XX
PS   Disclosure; SEQ ID NO 11910; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 7202 BP; 1552 A; 1818 C; 1967 G; 1865 T; 0 other;
```

```
  Query Match             7.8%;  Score 77.6;  DB 22;  Length 7202;
  Best Local Similarity   47.9%;  Pred. No. 5.8e-06;
  Matches  224;  Conservative    0;  Mismatches  244;  Indels    0;  Gaps    0;

Qy        366 GGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 425
              |||  | ||| | | | | ||||| |   | |||   | | | | | |    ||   | |
Db       1447 GGGTGGGTGGATGGGTGGGAGGGCAGAAGGATGGATAGATGGGTAGATGAGAGGATGGAT 1506

Qy        426 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 485
              | | |   | |||   | | |   |   |  ||   |   | | | | | |||   | |
Db       1507 GGATGGAAAGATGGGTGGATGGGAGGGAGGGAGGGAAGGAAGATGGGTGGGTGGATGGAT 1566

Qy        486 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 545
              | | | | | |||   |   |   | | | |||   | | | | | | | |||   | |
Db       1567 GGATGGATAGATGGAAGGGCAGGAAGATGGGTGGGGGGGTGGGTGGATGGATGGATGGAT 1626

Qy        546 TGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCT 605
              | ||   ||  || |||   | |     | |    ||  || | ||  |   |   |
Db       1627 CGATGGATGGCTAGAAGGAAGGATGAGAAGATGGATGGGAGGGAGGCAGGGAGGGAAGGA 1686

Qy        606 TGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGC 665
              |||||  |||  | |||| | |||  ||   | |||   |   | |||||  |   |
Db       1687 AGATAGGTGGGTGGGTGGGTGGGTAGATGGATGGATGGATAGAAGGAAGGGCCGGAAGAT 1746

Qy        666 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGC 725
              |  ||  |||  | ||||  | | |  |||    |||   | | |  ||   | |||
Db       1747 GGGCAGGTGGGTGGGTGGGTGGTTGGGTGGGTGAATGGATGGATGGATGGATGGATGGAT 1806

Qy        726 TGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGC 785
              | | |  ||     |||  ||| |  || | |||||  ||| |  ||  ||||
Db       1807 GGATGGATGGATGGATGGATGGCTGGCTGGCCGGCTGGCTGGCTGGCTGGCTGTCT 1866

Qy        786 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCT 833
              |||    ||   ||||    | | | |||||  |   | ||| | | |
Db       1867 GGCTGTCTGGCTGGCTGTAAAGATGGAAGGGCAGGAAGGTGGATGGGT 1914
```

```
RESULT 32
AAK71750
ID   AAK71750 standard; DNA; 7202 BP.
XX
AC   AAK71750;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:26562.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
```

PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 26562; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the

```
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 7202 BP; 1552 A; 1818 C; 1967 G; 1865 T; 0 other;

  Query Match             7.8%;  Score 77.6;  DB 22;  Length 7202;
  Best Local Similarity   47.9%;  Pred. No. 5.8e-06;
  Matches  224;  Conservative    0;  Mismatches  244;  Indels    0;  Gaps    0;

Qy        366 GGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 425
Db       1447 GGGTGGGTGGATGGGTGGGAGGGCAGAAGGATGGATAGATGGGTAGATGAGAGGATGGAT 1506

Qy        426 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 485
Db       1507 GGATGGAAAGATGGGTGGATGGGAGGGAGGGAGGGAAGGAAGATGGGTGGGTGGATGGAT 1566

Qy        486 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 545
Db       1567 GGATGGATAGATGGAAGGGCAGGAAGATGGGTGGGGGGGTGGGTGGATGGATGGATGGAT 1626

Qy        546 TGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCT 605
Db       1627 CGATGGATGGCTAGAAGGAAGGATGAGAAGATGGATGGGAGGGAGGCAGGGAGGGAAGGA 1686

Qy        606 TGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGC 665
Db       1687 AGATAGGTGGGTGGGTGGGTGGGTAGATGGATGGATGGATAGAAGGAAGGGCCGGAAGAT 1746

Qy        666 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGC 725
Db       1747 GGGCAGGTGGGTGGGTGGGTGGTTGGGTGGGTGAATGGATGGATGGATGGATGGATGGAT 1806

Qy        726 TGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGC 785
Db       1807 GGATGGATGGATGGATGGATGGCTGGCTGGCCGGCTGGCTGGCTGGCTGGCTGGCTGTCT 1866

Qy        786 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCT 833
Db       1867 GGCTGTCTGGCTGGCTGTAAAGATGGAAGGGCAGGAAGGTGGATGGGT 1914


RESULT 33
ABS65032
ID   ABS65032 standard; DNA; 42999 BP.
XX
AC   ABS65032;
XX
DT   15-NOV-2002  (first entry)
XX
DE   Invertebrate foraging behaviour associated human DNA sequence #12.
XX
KW   Attention deficit hyperactivity disorder; ADHD; hypertension;
KW   invertebrate foraging behaviour; nitric oxide; hypotensive;
KW   cGMP-dependent protein kinase; human; neuroleptic; ds;
KW   gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200259370-A2.
XX
PD   01-AUG-2002.
XX
PF   13-DEC-2001; 2001WO-US48087.
XX
PR   15-DEC-2000; 2000US-0738630.
XX
PA   (NEUR-) NEUROSCIENCES RES FOUND INC.
XX
PI   Greenspan RJ,  Shaw PJ;
XX
DR   WPI; 2002-636544/68.
XX
PT   Identifying a compound that modulates an attention deficit
PT   hyperactivity disorder (ADHD) for treating e.g. ADHD or hypertension,
PT   comprises measuring a foraging behaviour an invertebrate -
XX
PS   Claim 89; Page 182-195; 246pp; English.
XX
CC   The present invention relates to a method of identifying a compound
CC   that modulates attention deficit hyperactivity disorder (ADHD) in
CC   a mammal. The method comprises administering a test compound to
CC   an invertebrate, and measuring a foraging behaviour of the
CC   invertebrate, where the compound that modulates the foraging behaviour
CC   of the invertebrate is characterised as a compound that modulates
CC   ADHD in a mammal. The compound identified by the new method of the
CC   invention is useful for diagnosing or treating ADHD, hypertension or
CC   other diseases associated with a nitric oxide/cGMP-dependent protein
CC   kinase network in a mammal, particularly humans. The method
CC   distinguishes a compound that has a specific effect on ADHD,
CC   hypertension or other diseases associated with a
CC   nitric oxide/cGMP-dependent kinase protein network
CC   in a mammal from a compound that has a non-specific effect. The
CC   present sequence represents a DNA sequence associated with
CC   invertebrate foraging behaviour.
XX
SQ   Sequence 42999 BP; 6411 A; 13605 C; 11491 G; 11479 T; 13 other;

  Query Match             7.7%;  Score 77;  DB 24;  Length 42999;
  Best Local Similarity   58.7%;  Pred. No. 7.9e-06;
  Matches  172;  Conservative    0;  Mismatches  115;  Indels    6;  Gaps    2;

Qy        263 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACA 322
Db      40286 GCTTTCTTGCTTGCTTGCTTTCGTGCTGTCTTGTTTCTCGATTTCTTTCTTTCTTTTGTT 40345

Qy        323 GTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGC-- 380
Db      40346 TCTTTCCTGCTTGCTTTCTTGCTTGATTGCTTTCGTGCTTTCTTGCTTTCTTGTTTTCTT 40405

Qy        381 ---TAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 437
Db      40406 TCTTTCTTTTGTTTCTTTCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCT 40465

Qy        438 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC-TTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
Db      40466 TTCGTGCTTTCTTGTTTTCTTGCTTTCTTTCTTTTGTTTCTTTCTTGCTTGCTTTCTTGC 40525

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
Db      40526 TTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTTCTTTCTTGCT 40578


RESULT 34
ABX11086
ID   ABX11086 standard; DNA; 42999 BP.
XX
AC   ABX11086;
XX
DT   24-APR-2003  (first entry)
XX
DE   Human ribosomal RNA (rRNA) gene.
XX
KW   Nucleic acid amplification; artificial chromosome isolation; MAC;
KW   DNA delivery; mammalian artificial chromosome; gene therapy; organ;
KW   humanised genetically transformed animal; chromosomal element;
KW   gene product production system; transgenic; centromere function;
KW   information storage vehicle; artificial chromosome vector; human;
KW   species-specific artificial chromosome; ribosomal RNA; rRNA; gene; ds.
XX
OS   Homo sapiens.
XX
```

PN US2002160410-A1.
XX
PD 31-OCT-2002.
XX
PF 17-APR-2002; 2002US-0125767.
XX
PR 28-NOV-2000; 2000US-0724693.
PR 10-APR-1997; 97US-0835682.
PR 10-APR-1996; 96US-0629822.
PR 15-JUL-1996; 96US-0682080.
PR 07-AUG-1996; 96US-0695191.
XX
PA (HADL/) HADLACZKY G.
PA (SZAL/) SZALAY A A.
XX
PI Hadlaczky G, Szalay AA;
XX
DR WPI; 2003-265757/26.
XX
PT Amplifying nucleic acid for constructing artificial chromosomes,
PT comprises introducing a DNA fragment into a cell to incorporate the DNA
PT into a chromosome, and identifying cells with chromosomes that has
PT undergone amplification -
XX
PS Disclosure; Page 76-94; 107pp; English.
XX
CC The present invention relates to methods for amplifying nucleic acids
CC in cells, for isolating artificial chromosomes, and preparing cell
CC lines that contain artificial chromosomes. The methods comprise
CC introducing a DNA fragment comprising a selectable marker into a cell,
CC growing the cell under selective conditions to produce cells that have
CC incorporated the DNA fragment or its portion into a chromosome, and
CC identifying from among the resulting cells those that include a
CC chromosome or its fragment with a portion that has undergone
CC amplification. The invention also discloses cells and cell lines that
CC contain the nucleic acids or artificial chromosomes. The methods of
CC the invention are useful for amplifying nucleic acids in cells, in
CC generating and isolating artificial chromosomes (e.g. mammalian
CC artificial chromosomes or MACs), and in delivering the chromosomes
CC to selected cells and tissues. The artificial chromosomes are useful
CC in gene therapy, gene product production systems, production
CC of humanised genetically transformed animal organs, production of
CC transgenic plants and animals that would employ chromosomal elements as
CC information storage vehicles, for analysis and study of centromere
CC function, for the production of artificial chromosome vectors, and for
CC the preparation of species-specific artificial chromosomes. The present
CC sequence representing the human ribosomal RNA (rRNA) gene is used in
CC the method of the present invention.
CC Note: All N nucleotides (EXCEPT those at positions 21526-21527, 23183,
CC 24975-24977, 28699, 28945, 34074 and 34474) have been added to the
CC sequence by the indexer to match the number of bases this sequence
CC is listed to contain in the sequence listing.
XX
SQ Sequence 42999 BP; 5970 A; 12784 C; 10787 G; 10765 T; 2693 other;

Query Match 7.5%; Score 75.4; DB 25; Length 42999;
Best Local Similarity 55.7%; Pred. No. 1.6e-05;
Matches 161; Conservative 0; Mismatches 126; Indels 2; Gaps 2;

```
Qy        263 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACA 322
              || | |||||| ||| || | | ||||| | |   |  ||  |
Db      40286 GCTTTCTTGCTTGCTTGCTTTCGTGCTGTCTNNNNTCTCGATTTCTTTCTTTCTTTTGTT 40345

Qy        323 GTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGC-T 381
              ||||||||||  || || || |||||     |       |  | || | || | | |
Db      40346 TCTTTCCTGCTTGCTTTCTTGCTTGATTGCTNNNNTGCTTTCTTGCTTTCTTGTTTTCTT 40405

Qy        382 AGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 441
                |    |  |  ||  | | ||||||| |    |  | ||||| |||||| ||| ||
Db      40406 TCTTTCTTTTGTTTCTTTCTTTCTTGCTTCCNNNNTTTCTTGCTTTCTTGCTTGCTTGCT 40465

Qy        442 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC-TGGCCTGCTTGCTTGCTCGCTGGC 500
              | | ||||| || | |  | ||||| ||| |    |  | | ||||||||||  || ||
Db      40466 TTCGTGCTTTCTTGTTTTCTTGCTTTCTTTCNNNNGTTTCTTTCTTGCTTGCTTTCTTGC 40525

Qy        501 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
                | |||| || || |||  | ||||||||||    |  || | ||| ||
Db      40526 TTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGNNNNCGTGCTTTCTTTCT 40574
```

RESULT 35
ABQ97101/c
ID ABQ97101 standard; cDNA; 428 BP.
XX
AC ABQ97101;
XX
DT 30-OCT-2002 (first entry)
XX
DE Mouse ES cell related cDNA SEQ ID NO 369.
XX
KW Mouse; ES cell; gene trapped sequence; GTS; gene expression;
KW development disorder; cell differentiation disorder; gene; ss.
XX
OS Mus sp.
XX
PN US2002081668-A1.
XX
PD 27-JUN-2002.
XX
PF 30-NOV-2000; 2000US-0728446.
XX
PR 20-NOV-1998; 98US-109302P.
PR 01-DEC-1999; 99US-168270P.
XX
PA (FRIE/) FRIEDRICH G.
PA (ZAMB/) ZAMBROWICZ B.
PA (SAND/) SANDS A T.
XX
PI Friedrich G, Zambrowicz B, Sands AT;
XX
DR WPI; 2002-626541/67.
XX
PT Novel murine polynucleotides that individually identify novel genes
PT into which a retroviral gene trap vector has been integrated, useful in
PT genomic analysis and in discovery, development of therapeutic and
PT diagnostic agents -
XX
PS Claim 2; SEQ ID NO 369; 29pp + Sequence Listing; English.
XX
CC The invention relates to isolated murine polynucleotides (I) comprising a
CC contiguous stretch of at least about 60 nucleotides of a sequence
CC (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences
CC (GTSs). The novel genes can be used in a process to identify novel
CC polynucleotide sequences by comparing them to the novel gene sequences.
CC The novel genes and cells are useful in functional genomic analysis and
CC in the discovery and development of new therapeutic and diagnostic agents
CC and methods. (I) is useful for identifying the coding regions of the
CC murine genome, to isolate cDNAs, genomic clones or full-length
CC genes/polynucleotides or homologues, heterologues, paralogues or
CC orthologues that are capable of hybridising to one or more of the GTSs
CC under stringent conditions. (I) can be incorporated into a phage display
CC system that can be used to screen for proteins or other ligands, that are
CC capable of binding an amino acid sequence encoded by an oligonucleotide
CC or polynucleotide sequence in at least one of the GTS sequences. (I) is
CC useful in arrays, such as gene chips, to identify and characterise
CC temporal and tissue specific gene expression, to identify the gene of
CC interest from many sources and for genetic manipulations such as
CC antisense inhibition and gene targeting. Decreasing the level of
CC expression of (I) and/or down regulating the activity of peptides or
CC proteins encoded by (I) is useful for treating development and cell
CC differentiation disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at sequdata.uspto.gov/sequence.html?DocID=20020081668.
XX
SQ Sequence 428 BP; 149 A; 158 C; 66 G; 55 T; 0 other;

```
  Query Match             7.5%;  Score 74.8;  DB 24;  Length 428;
  Best Local Similarity   61.1%;  Pred. No. 1.8e-05;
  Matches  121;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy        377 TTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 436
              ||||| |  ||   | | |||||  |||| | | ||| | ||||  | ||||| | | ||
Db        218 TTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGC 159

Qy        437 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
              |||   ||| | ||| | | ||||  | | ||| | | |||||  |||| | | ||| |
Db        158 TGGATGGCTGGGTTGGTGGATGGCTGGATGGCTGGATGGCTGGTTTGCTGGATGGCTGGA 99

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 556
              ||||  | ||| | | | |||||   ||| | ||| | | | |   | ||||||  |
Db         98 TGGCTGGGTTGGTGGATGGCTGGATGGCTGGGTTGGTGGATTGTTGGATTGCTGGTGACT 39

Qy        557 CAGTAGGGCTGCTTGCTT 574
                ||  | |  ||| ||
Db         38 GGTTATTGGTTGTTGGTT 21


RESULT 36
ABQ97086/c
ID   ABQ97086 standard; cDNA; 436 BP.
XX
AC   ABQ97086;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Mouse ES cell related cDNA SEQ ID NO 354.
XX
KW   Mouse; ES cell; gene trapped sequence; GTS; gene expression;
KW   development disorder; cell differentiation disorder; gene; ss.
XX
OS   Mus sp.
XX
PN   US2002081668-A1.
XX
PD   27-JUN-2002.
XX
PF   30-NOV-2000; 2000US-0728446.
XX
PR   20-NOV-1998;   98US-109302P.
PR   01-DEC-1999;   99US-168270P.
XX
PA   (FRIE/) FRIEDRICH G.
PA   (ZAMB/) ZAMBROWICZ B.
PA   (SAND/) SANDS A T.
XX
PI   Friedrich G,  Zambrowicz B,  Sands AT;
XX
DR   WPI; 2002-626541/67.
XX
PT   Novel murine polynucleotides that individually identify novel genes
PT   into which a retroviral gene trap vector has been integrated, useful in
PT   genomic analysis and in discovery, development of therapeutic and
PT   diagnostic agents -
XX
PS   Claim 2; SEQ ID NO 354; 29pp + Sequence Listing; English.
XX
CC   The invention relates to isolated murine polynucleotides (I) comprising a
CC   contiguous stretch of at least about 60 nucleotides of a sequence
CC   (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences
CC   (GTSs). The novel genes can be used in a process to identify novel
CC   polynucleotide sequences by comparing them to the novel gene sequences.
CC   The novel genes and cells are useful in functional genomic analysis and
CC   in the discovery and development of new therapeutic and diagnostic agents
CC   and methods. (I) is useful for identifying the coding regions of the
CC   murine genome, to isolate cDNAs, genomic clones or full-length
CC   genes/polynucleotides or homologues, heterologues, paralogues or
CC   orthologues that are capable of hybridising to one or more of the GTSs
CC   under stringent conditions. (I) can be incorporated into a phage display
CC   system that can be used to screen for proteins or other ligands, that are
CC   capable of binding an amino acid sequence encoded by an oligonucleotide
CC   or polynucleotide sequence in at least one of the GTS sequences. (I) is
CC   useful in arrays, such as gene chips, to identify and characterise
CC   temporal and tissue specific gene expression, to identify the gene of
CC   interest from many sources and for genetic manipulations such as
CC   antisense inhibition and gene targeting. Decreasing the level of
CC   expression of (I) and/or down regulating the activity of peptides or
CC   proteins encoded by (I) is useful for treating development and cell
CC   differentiation disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at sequdata.uspto.gov/sequence.html?DocID=20020081668.
XX
SQ   Sequence 436 BP; 150 A; 157 C; 68 G; 60 T; 1 other;

  Query Match             7.5%;  Score 74.8;  DB 24;  Length 436;
  Best Local Similarity   61.1%;  Pred. No. 1.8e-05;
  Matches  121;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy        377 TTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 436
              ||||| |  ||   | | |||||  |||| | | ||| | ||||  | ||||| | | ||
Db        226 TTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGC 167

Qy        437 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
              |||   ||| | ||| | | ||||  | | ||| | | |||||  |||| | | ||| |
Db        166 TGGATGGCTGGGTTGGTGGATGGCTGGATGGCTGGATGGCTGGTTTGCTGGATGGCTGGA 107

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 556
              ||||  | ||| | | | |||||   ||| | ||| | | | |   | ||||||  |
Db        106 TGGCTGGGTTGGTGGATGGCTGGATGGCTGGGTTGGTGGATTGTTGGATTGCTGGTGACT 47

Qy        557 CAGTAGGGCTGCTTGCTT 574
                ||  | |  ||| ||
Db         46 GGTTATTGGTTGTTGGTT 29


RESULT 37
ABA19578
ID   ABA19578 standard; DNA; 24512 BP.
XX
AC   ABA19578;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 11909.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
```

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 11909; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone

```
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 24512 BP; 5908 A; 6031 C; 6107 G; 6466 T; 0 other;

  Query Match             7.4%;  Score 74.2;  DB 22;  Length 24512;
  Best Local Similarity   48.3%;  Pred. No. 2.6e-05;
  Matches  208;  Conservative    0;  Mismatches  223;  Indels    0;  Gaps    0;

Qy        415 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 474
              | |||   | | | | | |  || | |   | | | | | |||   | |     | |
Db      18843 GATGGGTGGGTGGATGGGTGGGAGGGCAGAAGGATGGATAGATGGGTAGATGAGAGGATG 18902

Qy        475 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 534
              | |||   | | | | |   | |||   | | |   |   |  ||   |  | | | |
Db      18903 GATGGATGGATGGATGGAAAGATGGGTGGATGGGAGGGAGGGAGGGAAGGAAGATGGGTG 18962

Qy        535 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCT 594
              | |||   | | | ||    ||  | |||| |   | | | |  ||||  ||   |
Db      18963 GGTGGATGGATGGATGGATAGATGGAAGGGCAGGAAGATGGGTGGGGGGGTGGGTGGATG 19022

Qy        595 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 654
              |   ||||| | ||| |  ||   |  ||   | |   | |   | ||||  |   | ||
Db      19023 GATGGATGGATCGATGGATGGCTAGAAGGAAGGATGAGAAGATGGATGGGAGGGAGGCAG 19082

Qy        655 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 714
              ||  |   ||  | |||  |||  | ||||  | |||  ||     |||   | | |
Db      19083 GGAGGGAAGGAAGATAGGTGGGTGGGTGGGTGGGTAGATGGATGGATGGATGGATGGATA 19142

Qy        715 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 774
              |   |  |||| |  ||  ||||   ||||  | | |  |    | |||   | | |  |
Db      19143 GAAGGAAGGGCCGGAAGATGGGCAGGTGGGTGGGTGGGTGAATGGATGGATGGATGGATG 19202

Qy        775 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTT 834
              |   | |||   | | |  ||   | |||   | | |  ||   ||  |||||| |||
Db      19203 GATGGATGGATGGATGGATGGATGGATGGATGGGTGGGTGGATGGCTGGCTGGCTGGCTG 19262

Qy        835 GCTTGCTTGCT 845
              ||| ||| |||
Db      19263 GCTGGCTGGCT 19273


RESULT 38
AAK67571
ID   AAK67571 standard; DNA; 24512 BP.
XX
AC   AAK67571;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:22383.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
```

PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 22383; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 24512 BP; 5908 A; 6031 C; 6107 G; 6466 T; 0 other;

Query Match 7.4%; Score 74.2; DB 22; Length 24512;
Best Local Similarity 48.3%; Pred. No. 2.6e-05;
Matches 208; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

```
Qy      415 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 474
            | |||   | | | | | | |  || | |   | | | | | |||   | |     | |
Db    18843 GATGGGTGGGTGGATGGGTGGGAGGGCAGAAGGATGGATAGATGGGTAGATGAGAGGATG 18902

Qy      475 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 534
            | |||   | | | | |   | |||   | | |   |   |  ||   |   | | | |
Db    18903 GATGGATGGATGGATGGAAAGATGGGTGGATGGGAGGGAGGGAGGGAAGGAAGATGGGTG 18962

Qy      535 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCT 594
            | |||   | | | ||    ||  | ||||| |   | | | |  ||||  ||   |
Db    18963 GGTGGATGGATGGATGGATAGATGGAAGGGCAGGAAGATGGGTGGGGGGGTGGGTGGATG 19022

Qy      595 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 654
            |   ||||| | ||| |  ||   |  ||   | |   | |   | ||||  |   | ||
Db    19023 GATGGATGGATCGATGGATGGCTAGAAGGAAGGATGAGAAGATGGATGGGAGGGAGGCAG 19082

Qy      655 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 714
            ||  |   ||  | |||  |||  | |||| | |||  ||     |||   | | |
Db    19083 GGAGGGAAGGAAGATAGGTGGGTGGGTGGGTGGGTAGATGGATGGATGGATGGATGGATA 19142

Qy      715 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 774
            |   |  |||| |  ||  ||||   ||||  | | |  |    | |||   | | |  |
Db    19143 GAAGGAAGGGCCGGAAGATGGGCAGGTGGGTGGGTGGGTGAATGGATGGATGGATGGATG 19202

Qy      775 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTT 834
            |   | |||   | | |  ||   | |||   | | |  ||   ||  ||||||| |||
Db    19203 GATGGATGGATGGATGGATGGATGGATGGATGGGTGGGTGGATGGCTGGCTGGCTGGCTG 19262

Qy      835 GCTTGCTTGCT 845
            ||| ||| |||
Db    19263 GCTGGCTGGCT 19273
```

RESULT 39
AAK71749
ID AAK71749 standard; DNA; 24512 BP.
XX
AC AAK71749;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:26561.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.

XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX

```
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 26561; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 24512 BP; 5908 A; 6031 C; 6107 G; 6466 T; 0 other;

  Query Match             7.4%;  Score 74.2;  DB 22;  Length 24512;
  Best Local Similarity   48.3%;  Pred. No. 2.6e-05;
  Matches  208;  Conservative    0;  Mismatches  223;  Indels    0;  Gaps    0;

Qy        415 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 474
              | |||   | | | | | | |  || | |   | | | | | |||   | |     | |
Db      18843 GATGGGTGGGTGGATGGGTGGGAGGGCAGAAGGATGGATAGATGGGTAGATGAGAGGATG 18902

Qy        475 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 534
              | |||   | | | | |   | |||   | | |   |   |  ||   |  | | | |
Db      18903 GATGGATGGATGGATGGAAAGATGGGTGGATGGGAGGGAGGGAGGGAAGGAAGATGGGTG 18962

Qy        535 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCT 594
              | |||   | | | ||    ||  | |||||  |   | | | |  ||||  ||   |
Db      18963 GGTGGATGGATGGATGGATAGATGGAAGGGCAGGAAGATGGGTGGGGGGGTGGGTGGATG 19022

Qy        595 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 654
              |   ||||| | ||| |  ||   |  ||   | |   | |   | ||||  |   | ||
Db      19023 GATGGATGGATCGATGGATGGCTAGAAGGAAGGATGAGAAGATGGATGGGAGGGAGGCAG 19082

Qy        655 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 714
              ||  |   ||  | ||| |||  | ||||  | |||  ||     |||   | | |
Db      19083 GGAGGGAAGGAAGATAGGTGGGTGGGTGGGTGGGTAGATGGATGGATGGATGGATA 19142

Qy        715 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 774
              |   |  |||| |  ||  ||||   |||| | | |  |    | |||  | | |  |
Db      19143 GAAGGAAGGGCCGGAAGATGGGCAGGTGGGTGGGTGGGTGAATGGATGGATGGATGGATG 19202

Qy        775 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTT 834
              |   | |||   | | |  ||   | |||   | | |  ||   ||  |||||||| |||
Db      19203 GATGGATGGATGGATGGATGGATGGATGGATGGGTGGGTGGATGGCTGGCTGGCTGGCTG 19262

Qy        835 GCTTGCTTGCT 845
              ||| ||| |||
Db      19263 GCTGGCTGGCT 19273


RESULT 40
ABQ97027/c
ID   ABQ97027 standard; cDNA; 388 BP.
XX
AC   ABQ97027;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Mouse ES cell related cDNA SEQ ID NO 295.
XX
KW   Mouse; ES cell; gene trapped sequence; GTS; gene expression;
KW   development disorder; cell differentiation disorder; gene; ss.
XX
OS   Mus sp.
XX
PN   US2002081668-A1.
XX
PD   27-JUN-2002.
XX
PF   30-NOV-2000; 2000US-0728446.
XX
PR   20-NOV-1998;   98US-109302P.
PR   01-DEC-1999;   99US-168270P.
XX
PA   (FRIE/) FRIEDRICH G.
PA   (ZAMB/) ZAMBROWICZ B.
PA   (SAND/) SANDS A T.
XX
PI   Friedrich G,  Zambrowicz B,  Sands AT;
XX
DR   WPI; 2002-626541/67.
XX
PT   Novel murine polynucleotides that individually identify novel genes
PT   into which a retroviral gene trap vector has been integrated, useful in
PT   genomic analysis and in discovery, development of therapeutic and
PT   diagnostic agents -
XX
PS   Claim 2; SEQ ID NO 295; 29pp + Sequence Listing; English.
XX
CC   The invention relates to isolated murine polynucleotides (I) comprising a
CC   contiguous stretch of at least about 60 nucleotides of a sequence
CC   (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences
CC   (GTSs). The novel genes can be used in a process to identify novel
CC   polynucleotide sequences by comparing them to the novel gene sequences.
CC   The novel genes and cells are useful in functional genomic analysis and
CC   in the discovery and development of new therapeutic and diagnostic agents
CC   and methods. (I) is useful for identifying the coding regions of the
CC   murine genome, to isolate cDNAs, genomic clones or full-length
CC   genes/polynucleotides or homologues, heterologues, paralogues or
CC   orthologues that are capable of hybridising to one or more of the GTSs
CC   under stringent conditions. (I) can be incorporated into a phage display
CC   system that can be used to screen for proteins or other ligands, that are
CC   capable of binding an amino acid sequence encoded by an oligonucleotide
CC   or polynucleotide sequence in at least one of the GTS sequences. (I) is
CC   useful in arrays, such as gene chips, to identify and characterise
CC   temporal and tissue specific gene expression, to identify the gene of
CC   interest from many sources and for genetic manipulations such as
CC   antisense inhibition and gene targeting. Decreasing the level of
CC   expression of (I) and/or down regulating the activity of peptides or
CC   proteins encoded by (I) is useful for treating development and cell
CC   differentiation disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at sequdata.uspto.gov/sequence.html?DocID=20020081668.
XX
SQ   Sequence 388 BP; 133 A; 140 C; 62 G; 51 T; 2 other;

  Query Match             7.3%;  Score 73.2;  DB 24;  Length 388;
  Best Local Similarity   63.8%;  Pred. No. 3.6e-05;
  Matches  111;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        377 TTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 436
              |||||  |  ||   | | |||||  |||| | | |||  | ||||  | |||||  | | ||
Db        177 TTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGC 118

Qy        437 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
              |||   ||| | ||| | | ||||  | | ||| | | |||||  |||| | | |||  |
Db        117 TGGATGGCTGGGTTGGTGGATGGCTGGATGGCTGGATGGCTGGTTTGCTGGATGGCTGGA 58
```

```
Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              ||||  | ||| | | | |||||   ||| | ||| | | | |   | ||||||
Db         57 TGGCTGGGTTGGTGGATGGCTGGATGGCTGGGTTGGTGGATTGTTGGATTGCTG 4
```

RESULT 41
AAI29504/c
ID AAI29504 standard; cDNA; 15720 BP.
XX
AC AAI29504;
XX
DT 12-OCT-2001 (first entry)
XX
DE C899P determined cDNA sequence.
XX
KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KW immunogenic; gene therapy; vaccine; colonic cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200149716-A2.
XX
PD 12-JUL-2001.
XX
PF 29-DEC-2000; 2000WO-US35596.
XX
PR 30-DEC-1999; 99US-0476296.
PR 10-JAN-2000; 2000US-0480321.
PR 15-FEB-2000; 2000US-0504629.
PR 06-MAR-2000; 2000US-0519444.
PR 19-MAY-2000; 2000US-0575251.
PR 29-JUN-2000; 2000US-0609448.
PR 28-AUG-2000; 2000US-0649811.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI King GE, Wang T, Jiang Y;
XX
DR WPI; 2001-441847/47.
XX
PT Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer -
XX
PS Claim 2; Page 426-431; 472pp; English.
XX
CC The present invention describes colon tumour associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon tumour associated protein (TCAP)
CC expression, such as colonic cancer. For example, (I) and (II) may be
CC used to treat disorders associated with decreased expression by
CC rectifying mutations or deletions in a patient's genome that affect the
CC activity of TCAPs by expressing inactive proteins or to supplement the
CC patients own production of them. Additionally, (II) may be used to
CC produce the TCAP proteins, by inserting the nucleic acids into a host
CC cell culturing the cell to express the protein. (II) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC polymerase chain reaction (PCR) and hybridisation assays to detect and
CC quantitate the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy. (I) may
CC also be used as antigens in the production of antibodies against TCAPs
CC and in assays to identify modulators of TCAP expression and activity.
CC Anti-(I) antibodies and antagonists may also be used to down regulate
CC TCAP expression and activity. The anti-(I) antibodies may also be used
CC as diagnostic agents for detecting the presence of TCAPs in samples
CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512
CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences
CC given in the exemplification of the present invention.
XX
SQ Sequence 15720 BP; 4107 A; 6921 C; 2862 G; 1830 T; 0 other;

Query Match 7.3%; Score 73.2; DB 22; Length 15720;
Best Local Similarity 45.3%; Pred. No. 4e-05;
Matches 302; Conservative 0; Mismatches 363; Indels 1; Gaps 1;

```
Qy        282 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 341
              || ||| ||  |||   | ||   || ||| ||| |||   ||     || ||| | |
Db       5254 TGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGAAGGGGTGGTCATGGTGGTTGGTGGAG 5195

Qy        342 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 401
              |||  |       ||  ||     |||  | ||   || |  | ||   ||| |  |
Db       5194 GGCTGGGAGTGGTAGTTGAGGAAGGGGTGGTTGTGGTGGTTATTGGAGAGCTGGGGGTGG 5135

Qy        402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461
              || ||| |      | ||| |     || ||| | |  |||  |       || |   |||
Db       5134 TGGTTGGTGAAGGGGTGGTCATGGTGGTTGTTGGAGGGCTGGGAGTGGTGGTTATTGGAG 5075

Qy        462 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 521
               ||| |       | | |   |- | | | |      |  |  | |   ||| |  |
Db       5074 AGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTATTGGAGGGCTGGGAGTGG 5015

Qy        522 TGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGG 581
              || ||| | |   |||||      || |  | |||  |  ||      || ||| |   |
Db       5014 TGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTGGTGGAG 4955

Qy        582 GGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTG-CTAGTAGGGCTGC 640
              ||  || |     ||  || |||   | | |||| |||| | ||  |  |  | | ||  |
Db       4954 GGCTGGGAGTGGTGGTTGGTGGAAGGGTGGTAGTGCTGGTTGGTGGAGTGATGGGTGTAG 4895

Qy        641 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCC 700
              | |   |  |  |||      || || | |  |||  | || | |  | || ||    |
Db       4894 TCGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTCGTTGTTGGAGGGC 4835

Qy        701 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 760
              ||||      |   ||  |  |||  | |     |  |  |||    || | || | ||
Db       4834 TGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGATTGTTGGAGGACTGGGAGTGGTGG 4775

Qy        761 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 820
              | |   |  |  |||      || || | |  |||  | || | || | | ||   ||
Db       4774 TTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTGGTGGAGGGC 4715

Qy        821 CTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTG 880
              |  |   || ||| | |   | |||| | || | |||   |     ||||      |
Db       4714 TGGGAGTGGTGGTTGGTGGAAGGGTAGTAGTGCTGGTTGGTGGAGTGATGGGCGTAGTCG 4655

Qy        881 CAGCTGGGAGAACAGAGTAGGGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAG 940
              |  ||      || |  ||   |  |    |  |     | ||  ||  ||||||  |
Db       4654 TTGGAGGACTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTGTTGGAGGGCTGG 4595

Qy        941 ACGAGG 946
              | ||
Db       4594 GAGTGG 4589
```

RESULT 42
ABL62358/c
ID ABL62358 standard; DNA; 15720 BP.
XX
AC ABL62358;
XX
DT 15-MAY-2002 (first entry)
XX
DE Colon adenocarcinoma related gene sequence SEQ ID NO:695.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.

XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
DR WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
PS Claim 1; SEQ ID 695; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ Sequence 15720 BP; 4107 A; 6921 C; 2862 G; 1830 T; 0 other;

Query Match 7.3%; Score 73.2; DB 24; Length 15720;
Best Local Similarity 45.3%; Pred. No. 4e-05;
Matches 302; Conservative 0; Mismatches 363; Indels 1; Gaps 1;

```
Qy       282 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 341
Db      5254 TGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGAAGGGGTGGTCATGGTGGTTGGTGGAG 5195

Qy       342 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 401
Db      5194 GGCTGGGAGTGGTAGTTGAGGAAGGGGTGGTTGTGGTGGTTATTGGAGAGCTGGGGGTGG 5135

Qy       402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461
Db      5134 TGGTTGGTGAAGGGGTGGTCATGGTGGTTGTTGGAGGGCTGGGAGTGGTGGTTATTGGAG 5075

Qy       462 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 521
Db      5074 AGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTATTGGAGGGCTGGGAGTGG 5015

Qy       522 TGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGG 581
Db      5014 TGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTGGTGGAG 4955

Qy       582 GGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTG-CTAGTAGGGCTGC 640
Db      4954 GGCTGGGAGTGGTGGTTGGTGGAAGGGTGGTAGTGCTGGTTGGTGGAGTGATGGGTGTAG 4895

Qy       641 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCC 700
Db      4894 TCGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTCGTTGTTGGAGGGC 4835

Qy       701 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 760
Db      4834 TGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGATTGTTGGAGGACTGGGAGTGGTGG 4775

Qy       761 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 820
Db      4774 TTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTGGTGGAGGGC 4715

Qy       821 CTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTG 880
Db      4714 TGGGAGTGGTGGTTGGTGGAAGGGTAGTAGTGCTGGTTGGTGGAGTGATGGGCGTAGTCG 4655

Qy       881 CAGCTGGGAGAACAGAGTAGGGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAG 940
Db      4654 TTGGAGGACTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTGTTGGAGGGCTGG 4595

Qy       941 ACGAGG 946
Db      4594 GAGTGG 4589
```

RESULT 43
ABZ33690/c
ID ABZ33690 standard; cDNA; 15720 BP.
XX
AC ABZ33690;
XX
DT 30-JAN-2003 (first entry)
XX
DE Human colon tumour cDNA for clone C899P SEQ ID NO:1058.
XX
KW Human; colon cancer; colon tumour; immunotherapy; diagnosis; cancer;
KW tumour; immune response; immunostimulant; cytostatic; vaccine;
KW gene; ss.
XX

OS Homo sapiens.
XX
PN WO200283070-A2.
XX
PD 24-OCT-2002.
XX
PF 09-APR-2002; 2002WO-US11475.
XX
PR 10-APR-2001; 2001US-0833263.
PR 03-AUG-2001; 2001US-0922217.
PR 19-DEC-2001; 2001US-0025380.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI Wang T, Jiang Y, Smith CL, King GE, Wang A, Clapper JD;
PI Skeiky YAW, Fanger GR, Vedvick TS, Carter D;
XX
DR WPI; 2003-067548/06.
DR P-PSDB; ABP55365.
XX
PT New polynucleotide, useful for the preparation of a composition for
PT stimulating an immune response against, or treating, cancer -
XX
PS Disclosure; Page 428-433; 537pp; English.
XX
CC The present invention describes compounds (I) for the immunotherapy and
CC diagnosis of colon cancer. Also described: (1) a method for detecting
CC the presence of cancer in a patient; (2) a method for stimulating and/or
CC expanding T cells specific for a tumour protein; (3) an isolated T cell
CC population comprising T cells prepared by the method of (2); (4) a method
CC for stimulating an immune response in a patient; (5) a method for
CC treating cancer in a patient; and (6) a method for inhibiting the
CC development of cancer in a patient. (I) have immunostimulant and
CC cytostatic activities and can be used in vaccines. ABZ32646 to ABZ33725
CC and ABP55343 to ABP55391 represent human colon cancer/tumour related
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 15720 BP; 4107 A; 6921 C; 2862 G; 1830 T; 0 other;

Query Match 7.3%; Score 73.2; DB 25; Length 15720;
Best Local Similarity 45.3%; Pred. No. 4e-05;
Matches 302; Conservative 0; Mismatches 363; Indels 1; Gaps 1;

```
Qy       282 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 341
            || ||| ||  |||   | ||   || ||| ||| |||   ||     || ||| | |
Db      5254 TGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGAAGGGGTGGTCATGGTGGTTGGTGGAG 5195

Qy       342 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 401
            |||  |       ||  ||     |||  | ||   || |  | ||   ||| |  |
Db      5194 GGCTGGGAGTGGTAGTTGAGGAAGGGGTGGTTGTGGTGGTTATTGGAGAGCTGGGGGTGG 5135

Qy       402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461
            || ||| |     | ||| |    || ||| | |  |||  |      || |   |||
Db      5134 TGGTTGGTGAAGGGGTGGTCATGGTGGTTGTTGGAGGGCTGGGAGTGGTGGTTATTGGAG 5075

Qy       462 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 521
             ||| |     | | |   |   | | | |     |  |  | |   ||| |  |
Db      5074 AGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTATTGGAGGGCTGGGAGTGG 5015

Qy       522 TGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGG 581
            || ||| | |   |||||      || |  | |||  |  ||      || ||| |   |
Db      5014 TGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTGGTGGAG 4955

Qy       582 GGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTG-CTAGTAGGGCTGC 640
            ||  || |    ||  || |||   | | |||| |||| | ||  |  |  | ||  |
Db      4954 GGCTGGGAGTGGTGGTTGGTGGAAGGGTGGTAGTGCTGGTTGGTGGAGTGATGGGTGTAG 4895

Qy       641 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCC 700
            | |   |   | |||      || || | |  |||  | || | |  | || ||    |
Db      4894 TCGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTCGTTGTTGGAGGGC 4835

Qy       701 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 760
```

```
             ||||      |  ||  | |||  | |     |  |  |||    || | || | ||
Db      4834 TGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGATTGTTGGAGGACTGGGAGTGGTGG 4775

Qy       761 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 820
             | |   |   | |||      || || | |  |||  | || | |||  ||  ||   ||
Db      4774 TTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTGGTGGAGGC 4715

Qy       821 CTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTG 880
               |  |   || ||| | |   | |||| | || | |||   |      ||||      |
Db      4714 TGGGAGTGGTGGTTGGTGGAAGGGTAGTAGTGCTGGTTGGTGGAGTGATGGGCGTAGTCG 4655

Qy       881 CAGCTGGGAGAACAGAGTAGGGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAG 940
               | ||      || | ||   |  |    | |     | ||  ||  ||||||  |
Db      4654 TTGGAGGACTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTGTTGGAGGGCTGG 4595

Qy       941 ACGAGG 946
               | ||
Db      4594 GAGTGG 4589
```

RESULT 44
ABQ97094/c
ID ABQ97094 standard; cDNA; 409 BP.
XX
AC ABQ97094;
XX
DT 30-OCT-2002 (first entry)
XX
DE Mouse ES cell related cDNA SEQ ID NO 362.
XX
KW Mouse; ES cell; gene trapped sequence; GTS; gene expression;
KW development disorder; cell differentiation disorder; gene; ss.
XX
OS Mus sp.
XX
PN US2002081668-A1.
XX
PD 27-JUN-2002.
XX
PF 30-NOV-2000; 2000US-0728446.
XX
PR 20-NOV-1998; 98US-109302P.
PR 01-DEC-1999; 99US-168270P.
XX
PA (FRIE/) FRIEDRICH G.
PA (ZAMB/) ZAMBROWICZ B.
PA (SAND/) SANDS A T.
XX
PI Friedrich G, Zambrowicz B, Sands AT;
XX
DR WPI; 2002-626541/67.
XX
PT Novel murine polynucleotides that individually identify novel genes
PT into which a retroviral gene trap vector has been integrated, useful in
PT genomic analysis and in discovery, development of therapeutic and
PT diagnostic agents -
XX
PS Claim 2; SEQ ID NO 362; 29pp + Sequence Listing; English.
XX
CC The invention relates to isolated murine polynucleotides (I) comprising a
CC contiguous stretch of at least about 60 nucleotides of a sequence
CC (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences
CC (GTSs). The novel genes can be used in a process to identify novel
CC polynucleotide sequences by comparing them to the novel gene sequences.
CC The novel genes and cells are useful in functional genomic analysis and
CC in the discovery and development of new therapeutic and diagnostic agents
CC and methods. (I) is useful for identifying the coding regions of the
CC murine genome, to isolate cDNAs, genomic clones or full-length
CC genes/polynucleotides or homologues, heterologues, paralogues or
CC orthologues that are capable of hybridising to one or more of the GTSs
CC under stringent conditions. (I) can be incorporated into a phage display
CC system that can be used to screen for proteins or other ligands, that are
CC capable of binding an amino acid sequence encoded by an oligonucleotide

```
CC   or polynucleotide sequence in at least one of the GTS sequences. (I) is
CC   useful in arrays, such as gene chips, to identify and characterise
CC   temporal and tissue specific gene expression, to identify the gene of
CC   interest from many sources and for genetic manipulations such as
CC   antisense inhibition and gene targeting. Decreasing the level of
CC   expression of (I) and/or down regulating the activity of peptides or
CC   proteins encoded by (I) is useful for treating development and cell
CC   differentiation disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at sequdata.uspto.gov/sequence.html?DocID=20020081668.
XX
SQ   Sequence 409 BP; 135 A; 153 C; 63 G; 53 T; 5 other;

  Query Match             7.2%;  Score 72.2;  DB 24;  Length 409;
  Best Local Similarity   63.2%;  Pred. No. 5.5e-05;
  Matches  110;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy        377 TTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 436
              ||||| |  ||   | | |||||  |||| | | ||| | ||||  | ||||| | | ||
Db        205 TTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGC 146

Qy        437 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
              |||   ||| | ||| | | ||||  | | ||| | | |||||  |||| | | ||| |
Db        145 TGGATGGCTGGGTTGGTGGATGGCTGGATGGCTGGATGGCTGGTTTGCTGGATGGCTGGA 86

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              ||||  | ||| | | | |||||   ||| | ||| | | | |   | ||| ||
Db         85 TGGCTGGGTTGGTGGATGGCTGGATGGCTGGGTTGGTGGATTGTTGGATTGNTG 32


RESULT 45
AAK69442
ID   AAK69442 standard; DNA; 16084 BP.
XX
AC   AAK69442;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24254.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
```

PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 24254; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX

SQ Sequence 16084 BP; 2716 A; 4761 C; 5154 G; 3453 T; 0 other;

Query Match 7.2%; Score 71.8; DB 22; Length 16084;
Best Local Similarity 49.9%; Pred. No. 7.3e-05;
Matches 211; Conservative 0; Mismatches 207; Indels 5; Gaps 1;

```
Qy       390 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 449
            ||  ||| ||   |  ||       |  |||||    ||| ||| |  |||    | |
Db      8290 TGAGCGCAGGAGCGTCTGGGCCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCTTGTCC 8349

Qy       450 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 509
             |    |  |||||    ||| ||| |  |||    |  | |    |  |||||    |||
Db      8350 TCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCTTGTCCTCACGGTGGGCCTATCAGCT 8409

Qy       510 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 569
              ||| |  |||      |||||  | | ||   |     | ||  | | | ||||| ||| .
Db      8410 GGCTGGAGGGC-----GGCTTGTCCTCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCT 8464

Qy       570 TGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCT 629
             ||    |   | || |     |  ||| |||  ||   | | ||      | |||||   |
Db      8465 TGTCCTCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCTTGTCCTCAGGGTGGGCCTAT 8524

Qy       630 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 689
              |||||  |||| ||| ||      | |||||   |     |||||  |||| |||
Db      8525 CAGCTGGCTGGAGGGCGGCTTGTCCTCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCT 8584

Qy       690 AGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCT 749
              ||        |||||   |     |||||  |||| ||| ||        |||||   |
Db      8585 TGTCCTCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCTTGTCCTCAGGGTGGGCCTAT 8644

Qy       750 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 809
              |||||  |||| ||| ||      | |||||   |     |||||  |||| |||
Db      8645 CAGCTGGCTGGAGGGCGGCTTGTCCTCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCT 8704

Qy       810 AGT 812
              ||
Db      8705 TGT 8707
```

RESULT 46
AAK82881/c
ID AAK82881 standard; DNA; 16084 BP.
XX
AC AAK82881;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:37693.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 37693; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting

```
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 16084 BP; 3453 A; 5154 C; 4761 G; 2716 T; 0 other;

  Query Match             7.2%;  Score 71.8;  DB 22;  Length 16084;
  Best Local Similarity   49.9%;  Pred. No. 7.3e-05;
  Matches  211;  Conservative    0;  Mismatches  207;  Indels    5;  Gaps    1;

Qy        390 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 449
              ||  ||| ||   |  ||      |  |||||    ||| ||| |  |||   |  |
Db       7795 TGAGCGCAGGAGCGTCTGGGCCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCTTGTCC 7736

Qy        450 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 509
              |    |  |||||    ||| ||| |  |||    |  | |    |  |||||    |||
Db       7735 TCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCTTGTCCTCACGGTGGGCCTATCAGCT 7676

Qy        510 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 569
               ||| |  |||      |||||  | | ||   |     | ||  | | | ||||| |||
Db       7675 GGCTGGAGGGC-----GGCTTGTCCTCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCT 7621

Qy        570 TGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCT 629
              ||    |  | || |    |  ||| |||  ||  | | ||      | |||||   |
Db       7620 TGTCCTCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCTTGTCCTCAGGGTGGGCCTAT 7561

Qy        630 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 689
              |||||  |||| ||| ||      | |||||   |     |||||  |||| |||
Db       7560 CAGCTGGCTGGAGGGCGGCTTGTCCTCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCT 7501

Qy        690 AGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCT 749
              ||        |||||   |     |||||  |||| ||| ||        |||||   |
Db       7500 TGTCCTCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCTTGTCCTCAGGGTGGGCCTAT 7441

Qy        750 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 809
              |||||  |||| ||| ||      | |||||   |     |||||  |||| |||
Db       7440 CAGCTGGCTGGAGGGCGGCTTGTCCTCAGGGTGGGCCTATCAGCTGGCTGGAGGGCGGCT 7381

Qy        810 AGT 812
              |||
Db       7380 TGT 7378


RESULT 47
ABQ97468/c
ID   ABQ97468 standard; cDNA; 392 BP.
XX
AC   ABQ97468;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Mouse ES cell related cDNA SEQ ID NO 736.
XX
KW   Mouse; ES cell; gene trapped sequence; GTS; gene expression;
KW   development disorder; cell differentiation disorder; gene; ss.
XX
OS   Mus sp.
XX
PN   US2002081668-A1.
XX
PD   27-JUN-2002.
XX
PF   30-NOV-2000; 2000US-0728446.
XX
PR   20-NOV-1998;   98US-109302P.
PR   01-DEC-1999;   99US-168270P.
XX
PA   (FRIE/) FRIEDRICH G.
PA   (ZAMB/) ZAMBROWICZ B.
PA   (SAND/) SANDS A T.
XX
PI   Friedrich G,  Zambrowicz B,  Sands AT;
XX
DR   WPI; 2002-626541/67.
XX
PT   Novel murine polynucleotides that individually identify novel genes
PT   into which a retroviral gene trap vector has been integrated, useful in
PT   genomic analysis and in discovery, development of therapeutic and
PT   diagnostic agents -
XX
PS   Claim 2; SEQ ID NO 736; 29pp + Sequence Listing; English.
XX
CC   The invention relates to isolated murine polynucleotides (I) comprising a
CC   contiguous stretch of at least about 60 nucleotides of a sequence
CC   (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences
CC   (GTSs). The novel genes can be used in a process to identify novel
CC   polynucleotide sequences by comparing them to the novel gene sequences.
CC   The novel genes and cells are useful in functional genomic analysis and
CC   in the discovery and development of new therapeutic and diagnostic agents
CC   and methods. (I) is useful for identifying the coding regions of the
CC   murine genome, to isolate cDNAs, genomic clones or full-length
CC   genes/polynucleotides or homologues, heterologues, paralogues or
CC   orthologues that are capable of hybridising to one or more of the GTSs
CC   under stringent conditions. (I) can be incorporated into a phage display
CC   system that can be used to screen for proteins or other ligands, that are
CC   capable of binding an amino acid sequence encoded by an oligonucleotide
CC   or polynucleotide sequence in at least one of the GTS sequences. (I) is
CC   useful in arrays, such as gene chips, to identify and characterise
CC   temporal and tissue specific gene expression, to identify the gene of
CC   interest from many sources and for genetic manipulations such as
CC   antisense inhibition and gene targeting. Decreasing the level of
CC   expression of (I) and/or down regulating the activity of peptides or
CC   proteins encoded by (I) is useful for treating development and cell
CC   differentiation disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at sequdata.uspto.gov/sequence.html?DocID=20020081668.
XX
SQ   Sequence 392 BP; 131 A; 143 C; 63 G; 52 T; 3 other;

  Query Match             7.1%;  Score 71.4;  DB 24;  Length 392;
  Best Local Similarity   62.4%;  Pred. No. 7.8e-05;
  Matches  111;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        397 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 456
              | ||  | | ||| | | |||||  |||| | | ||| | ||||  | ||||| | | ||
Db        182 TTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGC 123

Qy        457 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 516
              |||   ||| | ||| | | ||||  | | ||| | | |||||  | || | | ||| |
Db        122 TGGATGGCTGGGTTGGTGGATGGCTGGATGGCTGGATGGCTGGTTTNCTGGATGGCTGGA 63

Qy        517 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 574
              ||||  | ||| | | | |||||   ||| | |  | |||  || ||   ||| | ||
Db         62 TGGCTGGGTTGGTGGATGGCTGGATGGCTGGGTTGGTGGATTGTTGGATAGCTGGTTT 5


RESULT 48
AAS46592
ID   AAS46592 standard; DNA; 6286 BP.
XX
AC   AAS46592;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #314.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW   cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
```

XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP02955.
XX
PR 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT analysing diseases associated with cytosine methylation state e.g.
PT cancer -
XX
PS Claim 1; SEQ ID No 314; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and
CC oncogenes having a sequence taken from 536 (actually 533 since
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC form part of a set of probes for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms and also to be used in an
CC array for analysing diseases associated with CpG dinucleotides e.g.
CC cancers and tumours. The probes can also be used in a method for
CC ascertaining genetic and/or epigenetic parameters for the diagnosis
CC and/or therapy of existing diseases or the predisposition to specific
CC diseases, by analysing cytosine methylations. The parameters may be
CC compared to another set of genetic and/or epigenetic parameters, the
CC differences serving as basis for diagnosis and/or prognosis events which
CC are disadvantageous to patients. The present sequence is one of the
CC 533 genomic sequences derived from tumour suppressor genes and
CC oncogenes. Sequences with even numbered Seq ID numbers are the
CC complementary sequence of the corresponding odd numbered sequence (e.g.
CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC is missing).
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6286 BP; 1526 A; 124 C; 1714 G; 2922 T; 0 other;

Query Match 7.1%; Score 71.4; DB 22; Length 6286;
Best Local Similarity 45.7%; Pred. No. 8.5e-05;
Matches 285; Conservative 0; Mismatches 336; Indels 2; Gaps 1;

```
Qy       181 AGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGC 240
Db      1958 ATGGATGGATGGATGGATGGATGTATAAGTGGATGGATGGATGGATGGATGGGTGA 2017

Qy       241 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 300
Db      2018 GTAGTAGGTGAGTGGGTAAATGGGTGGGTGGATGGGTGGATGGATGGATGTATAGATGGA 2077

Qy       301 AGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 360
Db      2078 TGGATGGATGGATGTATAGGTGGATGGATGGATGGATGGATGGATGGATGTATAAG 2137

Qy       361 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 420
```

```
Db      2138 TGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGTATAGATGGA 2197

Qy       421 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 480
Db      2198 TGGATGGATGGATGGATGGATGGATGGATGTATAAATGGATGGATGGATGGATGGATGGA 2257

Qy       481 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 540
Db      2258 TGGATGGATGGATGGATGGATGGATGTATAGGTGGATGGATGGATGGATGGATGGATGGA 2317

Qy       541 CTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGA 600
Db      2318 TGGATGGATGGATGTATAAGTGGATGGATGGATGTATAAGTGGATGGATGGATGGATGGA 2377

Qy       601 TGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 660
Db      2378 TGGATGGATGGATGGATGGATGGGTGAGTAGTAGG--TGAGTGGGTAAATGGGTGGGTGG 2435

Qy       661 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 720
Db      2436 ATGGGTGGATGGATGGATGTATAGATGGATGGACGTATGGATGTATAGGTGGATGGATGG 2495

Qy       721 TGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 780
Db      2496 ATGGATGGATGAGTAGTAGGTGAGTGGGTAAATGGGTGGGTGGATTGGTAGATGGATGGA 2555

Qy       781 TGGGCTGCTAGTAGGGCTGCTGG 803
Db      2556 TTGAATGTTGGAGGAATTGATGG 2578
```

RESULT 49
AAS70903/c
ID AAS70903 standard; cDNA; 3588 BP.
XX
AC AAS70903;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #6707.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG06716. .
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 6707; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The

```
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3588 BP; 930 A; 1100 C; 709 G; 849 T; 0 other;

  Query Match             7.1%;  Score 71.2;  DB 23;  Length 3588;
  Best Local Similarity   45.7%;  Pred. No. 9.1e-05;
  Matches  326;  Conservative    0;  Mismatches  383;  Indels    5;  Gaps    2;

Qy        175 GGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTG-CCTGCT 233
              ||| |||| || | | || |||| |||| | |  || ||| | ||||
Db       1784 GGGTGCAGGTGATGCTGATTGGGTGCAGGTGATGCTGGCTGATTGGGTGCAGGTGATGCT 1725

Qy        234 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGG 293
               ||||  | |  | |  | |  | ||| |  | ||  ||| | |
Db       1724 GGCTGATTGGGTGCAGGTGTTAGCTGATTGGGTGCAGGTGATGCTGATTGGGTGCAGGTG 1665

Qy        294 GGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCT 353
              | |  | |||  |||  |  | |  | ||  |||  |  ||| ||
Db       1664 ATGCTGGCTGACTGGATGCAAGTGATGCTGGCCGATTGGGTGCAGGTGATGCTGGATGAC 1605

Qy        354 GAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCT 413
              || |  ||  |||  | ||  | |||  | |  |  ||  ||| |
Db       1604 TGGGTGCAGGTGATGCTAGATGACTGGGTGCAGGCGATGTTGGCTGATTGGGTGCAGGTG 1545

Qy        414 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 473
              | | ||| |||  |||  |  |||  ||  |||  |  ||| |||  |||
Db       1544 ATGTTGGCTGATTGGGTGCAGGTGATGCTGGCTGTGTGCAGGTGATGCTGATTGGGTGCA 1485

Qy        474 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 533
              |  |  ||| |||  ||  || ||  || |||  || ||  || |||
Db       1484 GGTGGTGCTGGTTGACTGGGTGCAGGTGCTGCTGATTGGGTGCAGGTGATGCTGATTGGA 1425

Qy        534 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCC 593
              || ||  || |  || |  ||  |||| ||| |||  || |  |||  |
Db       1424 TGCAGGTGATGCTGATTGGGTGCAGGTGATGCTGGTTGATTG----GGTGCAGGTGATGC 1369

Qy        594 TGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 653
              || ||||  |  ||  ||  |||| | || |||  ||  |||| | || |||  ||
Db       1368 TGGCTGACTGGATGCAGGTGATGCTGATTGGGTGCAGGTGATGCTGATTGGATGCAGGTG 1309

Qy        654 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTA 713
              ||||  | || |||  ||  ||||  |  ||  |  ||  |  ||||| | |
Db       1308 ATGCTGATTGGGTGCAGGTGATGCTGGCTGATTGGGTGCGGGTGATGCTGGCTGATTGGG 1249

Qy        714 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 773
              |  ||||  |  || | |  |  |  ||  |  |  |  |  |
Db       1248 TGCAGATGATGCTGGCTGATTGGGTGCAAGTGATGCTGATTGATTGGGTGCAGGTGATGC 1189

Qy        774 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCT 833
              ||||| | || |||  |  ||||  |  ||  || ||  |||  | |||
Db       1188 TGGCTGATTGGGTGCAGATGATGCTGGCTGATTGGGTGTGGGTGATGCTGGCTGATTGGG 1129

Qy        834 TGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGG 887
              |||  |  |  | | |  ||| |  |  |  |||| |  || ||
Db       1128 TGCAGGTGATGCTGGCTGACTGGGTGCAGGTGATGCTGATTGCTTGGGTGCAGG 1075
```

```
RESULT 50
AAS92599/c
ID   AAS92599 standard; cDNA; 3588 BP.
XX
AC   AAS92599;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #28403.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG28412.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 28403; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3588 BP; 930 A; 1100 C; 709 G; 849 T; 0 other;

  Query Match             7.1%;  Score 71.2;  DB 23;  Length 3588;
  Best Local Similarity   45.7%;  Pred. No. 9.1e-05;
  Matches  326;  Conservative    0;  Mismatches  383;  Indels    5;  Gaps    2;

Qy        175 GGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTG-CCTGCT 233
              ||| |||| || | | || |||| |||| | |  || ||| | ||||
Db       1784 GGGTGCAGGTGATGCTGATTGGGTGCAGGTGATGCTGGCTGATTGGGTGCAGGTGATGCT 1725

Qy        234 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGG 293
               ||||  | |  | |  | |  | ||| |  | ||  ||| | |
Db       1724 GGCTGATTGGGTGCAGGTGTTAGCTGATTGGGTGCAGGTGATGCTGATTGGGTGCAGGTG 1665
```

```
Qy        294 GGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCT 353
              | |  | |||  |||    |  | |   |  ||  |||  |     |||  ||
Db       1664 ATGCTGGCTGACTGGATGCAAGTGATGCTGGCCGATTGGGTGCAGGTGATGCTGGATGAC 1605

Qy        354 GAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCT 413
              || |   ||   |||    |  ||   | |||     | |   |  ||  |||  |
Db       1604 TGGGTGCAGGTGATGCTAGATGACTGGGTGCAGGCGATGTTGGCTGATTGGGTGCAGGTG 1545

Qy        414 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 473
              | | ||| |||  |||  |     |||  ||  |||  |     ||| |||  |||
Db       1544 ATGTTGGCTGATTGGGTGCAGGTGATGCTGGCTGTGTGCAGGTGATGCTGATTGGGTGCA 1485

Qy        474 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 533
              |  |  ||| |||  ||   || ||      || |||   || ||      || |||
Db       1484 GGTGGTGCTGGTTGACTGGGTGCAGGTGCTGCTGATTGGGTGCAGGTGATGCTGATTGGA 1425

Qy        534 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCC 593
              || ||       || |  || |   ||   |||| ||| |||    || |  |||   |
Db       1424 TGCAGGTGATGCTGATTGGGTGCAGGTGATGCTGGTTGATTG----GGTGCAGGTGATGC 1369

Qy        594 TGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 653
              || ||||  |  ||   ||   |||| | || |||  ||   |||| | || |||  ||
Db       1368 TGGCTGACTGGATGCAGGTGATGCTGATTGGGTGCAGGTGATGCTGATTGGATGCAGGTG 1309

Qy        654 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTA 713
              ||||  | || |||  ||   ||||   |  ||   |  ||       |||||  | |
Db       1308 ATGCTGATTGGGTGCAGGTGATGCTGGCTGATTGGGTGCGGGTGATGCTGGCTGATTGGG 1249

Qy        714 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 773
              |       ||||   |   || | |  |       |   |  ||   |   |  |
Db       1248 TGCAGATGATGCTGGCTGATTGGGTGCAAGTGATGCTGATTGATTGGGTGCAGGTGATGC 1189

Qy        774 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCT 833
              |||||  | || |||   |   ||||   |  ||   || ||     |||   |  |||
Db       1188 TGGCTGATTGGGTGCAGATGATGCTGGCTGATTGGGTGTGGGTGATGCTGGCTGATTGGG 1129

Qy        834 TGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGG 887
              |||  |       |   | | |  |||  |  |    |  ||||  |  || ||
Db       1128 TGCAGGTGATGCTGGCTGACTGGGTGCAGGTGATGCTGATTGCTTGGGTGCAGG 1075
```

RESULT 51
AAK85846
ID AAK85846 standard; DNA; 7614 BP.
XX
AC AAK85846;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40658.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 40658; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 7614 BP; 2213 A; 1415 C; 1685 G; 2301 T; 0 other;

Query Match 7.1%; Score 71.2; DB 22; Length 7614;
Best Local Similarity 45.7%; Pred. No. 9.3e-05;
Matches 326; Conservative 0; Mismatches 383; Indels 5; Gaps 2;

```
Qy       175 GGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTG-CCTGCT 233
            ||| |||| || | | || |||| |||| |     |  || ||| |  ||||
Db      3189 GGGTGCAGGTGATGCTGATTGGGTGCAGGTGATGCTGGCTGATTGGGTGCAGGTGATGCT 3248

Qy       234 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGG 293
             ||||  | |  | |  | |  | ||| |      |  || ||| | |
Db      3249 GGCTGATTGGGTGCAGGTGTTAGCTGATTGGGTGCAGGTGATGCTGATTGGGTGCAGGTG 3308

Qy       294 GGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCT 353
              | |  | ||| |||    |  | |  |  || ||| |     ||| ||
Db      3309 ATGCTGGCTGACTGGATGCAAGTGATGCTGGCCGATTGGGTGCAGGTGATGCTGGATGAC 3368

Qy       354 GAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCT 413
             ||  |  ||  |||  | ||  | |||     | |  |  || ||| |
Db      3369 TGGGTGCAGGTGATGCTAGATGACTGGGTGCAGGCGATGTTGGCTGATTGGGTGCAGGTG 3428

Qy       414 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 473
             | | ||| |||  |||  |     |||  ||  |||  |     ||| |||  |||
Db      3429 ATGTTGGCTGATTGGGTGCAGGTGATGCTGGCTGTGTGCAGGTGATGCTGATTGGGTGCA 3488

Qy       474 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 533
             |  |  ||| |||  ||   || ||      || |||   || ||       || |||
Db      3489 GGTGGTGCTGGTTGACTGGGTGCAGGTGCTGCTGATTGGGTGCAGGTGATGCTGATTGGA 3548

Qy       534 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCC 593
             || ||      || |  || |   ||   |||| ||| |||    || |  |||    |
Db      3549 TGCAGGTGATGCTGATTGGGTGCAGGTGATGCTGGTTGATTG----GGTGCAGGTGATGC 3604

Qy       594 TGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 653
             || ||||  |  ||   ||  |||| | || |||  ||   |||| | || ||| |
Db      3605 TGGCTGACTGGATGCAGGTGATGCTGATTGGGTGCAGGTGATGCTGATTGGATGCAGGTG 3664

Qy       654 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTA 713
             ||||  | ||  |||  ||   ||||   |  ||   |  ||       ||||| | |
Db      3665 ATGCTGATTGGGTGCAGGTGATGCTGGCTGATTGGGTGCGGGTGATGCTGGCTGATTGGG 3724

Qy       714 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 773
             |        ||||   |  || | |  |        |   |  ||   |  |  |
Db      3725 TGCAGATGATGCTGGCTGATTGGGTGCAAGTGATGCTGATTGATTGGGTGCAGGTGATGC 3784

Qy       774 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCT 833
             |||||  | ||  |||   |   ||||   |  ||   || ||     |||  | |||
Db      3785 TGGCTGATTGGGTGCAGATGATGCTGGCTGATTGGGTGTGGGTGATGCTGGCTGATTGGG 3844

Qy       834 TGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGG 887
             |||  |      |   | | |  |||  |  |    |  ||||  |  || ||
Db      3845 TGCAGGTGATGCTGGCTGACTGGGTGCAGGTGATGCTGATTGCTTGGGTGCAGG 3898
```

RESULT 52
AAL40272/c
ID AAL40272 standard; DNA; 157 BP.
XX
AC AAL40272;
XX
DT 13-SEP-2002 (first entry)

```
XX
DE   BASB204 related DNA sequence SEQ ID No 33.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   P-PSDB; AAO21919.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear -
XX
PS   Example 2; Fig 3; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 157 BP; 66 A; 33 C; 28 G; 30 T; 0 other;

  Query Match             7.1%;  Score 70.8;  DB 24;  Length 157;
  Best Local Similarity   86.7%;  Pred. No. 9.9e-05;
  Matches   78;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        401 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 460
              ||||||||||||| ||| || ||||||||||||| ||| || ||||||||||||| ||| ||
Db        100 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 41

Qy        461 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              ||||||||||||| ||| || |||||||||
Db        '40 TTGCTTGCTTGCTTGCTTGCTTGCTT 11
```

```
RESULT 53
ABQ96978/c
ID   ABQ96978 standard; cDNA; 379 BP.
XX
AC   ABQ96978;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Mouse ES cell related cDNA SEQ ID NO 246.
XX
KW   Mouse; ES cell; gene trapped sequence; GTS; gene expression;
KW   development disorder; cell differentiation disorder; gene; ss.
XX
OS   Mus sp.
XX
PN   US2002081668-A1.
XX
PD   27-JUN-2002.
XX
PF   30-NOV-2000; 2000US-0728446.
XX
PR   20-NOV-1998;   98US-109302P.
PR   01-DEC-1999;   99US-168270P.
XX
PA   (FRIE/) FRIEDRICH G.
PA   (ZAMB/) ZAMBROWICZ B.
PA   (SAND/) SANDS A T.
XX
PI   Friedrich G,  Zambrowicz B,  Sands AT;
XX
DR   WPI; 2002-626541/67.
XX
PT   Novel murine polynucleotides that individually identify novel genes
PT   into which a retroviral gene trap vector has been integrated, useful in
PT   genomic analysis and in discovery, development of therapeutic and
PT   diagnostic agents -
XX
PS   Claim 2; SEQ ID NO 246; 29pp + Sequence Listing; English.
XX
CC   The invention relates to isolated murine polynucleotides (I) comprising a
CC   contiguous stretch of at least about 60 nucleotides of a sequence
CC   (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences
CC   (GTSs). The novel genes can be used in a process to identify novel
CC   polynucleotide sequences by comparing them to the novel gene sequences.
CC   The novel genes and cells are useful in functional genomic analysis and
CC   in the discovery and development of new therapeutic and diagnostic agents
CC   and methods. (I) is useful for identifying the coding regions of the
CC   murine genome, to isolate cDNAs, genomic clones or full-length
CC   genes/polynucleotides or homologues, heterologues, paralogues or
CC   orthologues that are capable of hybridising to one or more of the GTSs
CC   under stringent conditions. (I) can be incorporated into a phage display
CC   system that can be used to screen for proteins or other ligands, that are
CC   capable of binding an amino acid sequence encoded by an oligonucleotide
CC   or polynucleotide sequence in at least one of the GTS sequences. (I) is
CC   useful in arrays, such as gene chips, to identify and characterise
CC   temporal and tissue specific gene expression, to identify the gene of
CC   interest from many sources and for genetic manipulations such as
CC   antisense inhibition and gene targeting. Decreasing the level of
CC   expression of (I) and/or down regulating the activity of peptides or
CC   proteins encoded by (I) is useful for treating development and cell
CC   differentiation disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at sequdata.uspto.gov/sequence.html?DocID=20020081668.
XX
SQ   Sequence 379 BP; 120 A; 143 C; 61 G; 50 T; 5 other;

  Query Match             7.1%;  Score 70.8;  DB 24;  Length 379;
  Best Local Similarity   59.1%;  Pred. No. 0.0001;
  Matches  117;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        377 TTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 436
              ||||| |  ||   | | |||||  |||| | | ||| | ||||  | ||||| | | ||
```

```
Db        198 TTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGC 139

Qy        437 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
              |||   ||| |  || |   || |    | ||| | | |||||  |||| | | ||| |
Db        138 TGGATGGCTGGGNTGGTGNATGNCTGNATGGCTGGATGGCTGGTTTGCTGGATGGCTGGA 79

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 556
              ||||  | ||| | | | |||||   ||| | ||| | | | |   | ||||||  |
Db         78 TGGCTGGGTTGGTGGATGGCTGGATGGCTGGGTTGGTGGATTGTTAGATTGCTGGTGACT 19

Qy        557 CAGTAGGGCTGCTTGCTT 574
                ||  | |  ||| ||
Db         18 GGTTATTGGTTGTTGGTT 1


RESULT 54
ABQ96982/c
ID   ABQ96982 standard; cDNA; 427 BP.
XX
AC   ABQ96982;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Mouse ES cell related cDNA SEQ ID NO 250.
XX
KW   Mouse; ES cell; gene trapped sequence; GTS; gene expression;
KW   development disorder; cell differentiation disorder; gene; ss.
XX
OS   Mus sp.
XX
PN   US2002081668-A1.
XX
PD   27-JUN-2002.
XX
PF   30-NOV-2000; 2000US-0728446.
XX
PR   20-NOV-1998;   98US-109302P.
PR   01-DEC-1999;   99US-168270P.
XX
PA   (FRIE/) FRIEDRICH G.
PA   (ZAMB/) ZAMBROWICZ B.
PA   (SAND/) SANDS A T.
XX
PI   Friedrich G,  Zambrowicz B,  Sands AT;
XX
DR   WPI; 2002-626541/67.
XX
PT   Novel murine polynucleotides that individually identify novel genes
PT   into which a retroviral gene trap vector has been integrated, useful in
PT   genomic analysis and in discovery, development of therapeutic and
PT   diagnostic agents  -
XX
PS   Claim 2; SEQ ID NO 250; 29pp + Sequence Listing; English.
XX
CC   The invention relates to isolated murine polynucleotides (I) comprising a
CC   contiguous stretch of at least about 60 nucleotides of a sequence
CC   (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences
CC   (GTSs). The novel genes can be used in a process to identify novel
CC   polynucleotide sequences by comparing them to the novel gene sequences.
CC   The novel genes and cells are useful in functional genomic analysis and
CC   in the discovery and development of new therapeutic and diagnostic agents
CC   and methods. (I) is useful for identifying the coding regions of the
CC   murine genome, to isolate cDNAs, genomic clones or full-length
CC   genes/polynucleotides or homologues, heterologues, paralogues or
CC   orthologues that are capable of hybridising to one or more of the GTSs
CC   under stringent conditions. (I) can be incorporated into a phage display
CC   system that can be used to screen for proteins or other ligands, that are
CC   capable of binding an amino acid sequence encoded by an oligonucleotide
CC   or polynucleotide sequence in at least one of the GTS sequences. (I) is
CC   useful in arrays, such as gene chips, to identify and characterise
CC   temporal and tissue specific gene expression, to identify the gene of
CC   interest from many sources and for genetic manipulations such as
CC   antisense inhibition and gene targeting. Decreasing the level of
CC   expression of (I) and/or down regulating the activity of peptides or
CC   proteins encoded by (I) is useful for treating development and cell
CC   differentiation disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at sequdata.uspto.gov/sequence.html?DocID=20020081668.
XX
SQ   Sequence 427 BP; 148 A; 155 C; 66 G; 54 T; 4 other;

  Query Match             7.1%;  Score 70.8;  DB 24;  Length 427;
  Best Local Similarity   59.4%;  Pred. No. 0.0001;
  Matches  117;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy        377 TTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 436
              ||||| |  ||   | |  ||||  |||| | | ||| | ||||  | |||||  | | ||
Db        217 TTGCTGGATGGCTGGATGNCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGC 158

Qy        437 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
              |||   ||| | ||| |   ||||  | | ||| | | |||||  ||||   | ||| |
Db        157 TGGATGGCTGGGTTGGTGNATGGCTGGATGGCTGGATGGCTGGTTTGCTGNATGGCTGGA 98

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 556
              ||||  | ||| | | | |||||   ||| | ||| | | | |   | ||||||  |
Db         97 TGGCTGGGTTGGTGGATGGCTGGATGGCTGGGTTGGTGGATTGTTGGATTGCTGGTGACT 38

Qy        557 CAGTAGGGCTGCTTGCT 573
                ||  | |  ||| |
Db         37 GGTTATTGGTTGTTGGT 21


RESULT 55
ABQ97057/c
ID   ABQ97057 standard; cDNA; 410 BP.
XX
AC   ABQ97057;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Mouse ES cell related cDNA SEQ ID NO 325.
XX
KW   Mouse; ES cell; gene trapped sequence; GTS; gene expression;
KW   development disorder; cell differentiation disorder; gene; ss.
XX
OS   Mus sp.
XX
PN   US2002081668-A1.
XX
PD   27-JUN-2002.
XX
PF   30-NOV-2000; 2000US-0728446.
XX
PR   20-NOV-1998;   98US-109302P.
PR   01-DEC-1999;   99US-168270P.
XX
PA   (FRIE/) FRIEDRICH G.
PA   (ZAMB/) ZAMBROWICZ B.
PA   (SAND/) SANDS A T.
XX
PI   Friedrich G,  Zambrowicz B,  Sands AT;
XX
DR   WPI; 2002-626541/67.
XX
PT   Novel murine polynucleotides that individually identify novel genes
PT   into which a retroviral gene trap vector has been integrated, useful in
PT   genomic analysis and in discovery, development of therapeutic and
PT   diagnostic agents  -
XX
PS   Claim 2; SEQ ID NO 325; 29pp + Sequence Listing; English.
XX
CC   The invention relates to isolated murine polynucleotides (I) comprising a
CC   contiguous stretch of at least about 60 nucleotides of a sequence
CC   (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences
CC   (GTSs). The novel genes can be used in a process to identify novel
```

```
CC   polynucleotide sequences by comparing them to the novel gene sequences.
CC   The novel genes and cells are useful in functional genomic analysis and
CC   in the discovery and development of new therapeutic and diagnostic agents
CC   and methods. (I) is useful for identifying the coding regions of the
CC   murine genome, to isolate cDNAs, genomic clones or full-length
CC   genes/polynucleotides or homologues, heterologues, paralogues or
CC   orthologues that are capable of hybridising to one or more of the GTSs
CC   under stringent conditions. (I) can be incorporated into a phage display
CC   system that can be used to screen for proteins or other ligands, that are
CC   capable of binding an amino acid sequence encoded by an oligonucleotide
CC   or polynucleotide sequence in at least one of the GTS sequences. (I) is
CC   useful in arrays, such as gene chips, to identify and characterise
CC   temporal and tissue specific gene expression, to identify the gene of
CC   interest from many sources and for genetic manipulations such as
CC   antisense inhibition and gene targeting. Decreasing the level of
CC   expression of (I) and/or down regulating the activity of peptides or
CC   proteins encoded by (I) is useful for treating development and cell
CC   differentiation disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at sequdata.uspto.gov/sequence.html?DocID=20020081668.
XX
SQ   Sequence 410 BP; 140 A; 151 C; 63 G; 55 T; 1 other;

  Query Match             7.1%;  Score 70.6;  DB 24;  Length 410;
  Best Local Similarity   62.6%;  Pred. No. 0.00011;
  Matches  109;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy        377 TTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 436
              ||||| |  ||   | | |||||  |||| | | ||| | ||||  | ||||| | | ||
Db        200 TTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGC 141

Qy        437 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
              |||   ||| | ||| | | ||||  | | ||| | | |||||  |||| | | ||| |
Db        140 TGGATGGCTGGGTTGGTGGATGGCTGGATGGCTGGATGGCTGGTTTGCTGGATGGCTGGA 81

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              ||||  | ||  | | | |||||   ||| | ||| | | | |   | | ||||
Db         80 TGGCTGGGTTCGTGGATGGCTGGATGGCTGGGTTGGTGGATTGTTGGATNGCTG 27


RESULT 56
AAL06216
ID   AAL06216 standard; DNA; 368 BP.
XX
AC   AAL06216;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 8904.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
```

PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; SEQ ID NO 8904; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 368 BP; 1 A; 99 C; 47 G; 221 T; 0 other;

Query Match 6.9%; Score 69.4; DB 22; Length 368;
Best Local Similarity 60.3%; Pred. No. 0.00019;
Matches 132; Conservative 0; Mismatches 86; Indels 1; Gaps 1;

```
Qy        325 AGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGT 384
              | || ||||| || ||| || |||||||   | |       |  | |  | || || |
Db         99 ATGCTTGCTTTCTTGCTTGCTTGCTTGCTTTCGTGCTTTCTTGCTTTCCCGTTTTCTTGC 158

Qy        385 AGGGCTGCTCGC-TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 443
                  | ||  | |    | ||| |||||| |||  | ||||||||||||  |  || |
Db        159 TTTCTTTCTTTCTTTTGTTTCTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTT 218

Qy        444 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 503
              |||| || |||  |    |  ||  ||  |  ||  |||||||||||| || ||| || ||
Db        219 CTTGTTTTCTCCTTTCTTTCTTTCTTTTGTTTCTTTCCTGCTTGCTTTCTTGCTTGCTTG 278

Qy        504 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 542
              ||| | ||||  || || | |||| || ||  ||  | |
Db        279 CTTTCGTGCTTTCTTGCTTTCTTGTTTTCTTTCTTTCTT 317
```

RESULT 57
AAL07359
ID AAL07359 standard; DNA; 368 BP.
XX
AC AAL07359;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 10047.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; SEQ ID NO 10047; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 368 BP; 1 A; 99 C; 47 G; 221 T; 0 other;

Query Match 6.9%; Score 69.4; DB 22; Length 368;
Best Local Similarity 60.3%; Pred. No. 0.00019;
Matches 132; Conservative 0; Mismatches 86; Indels 1; Gaps 1;

```
Qy        325 AGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGT 384
              | || ||||| || ||| || ||||||||   | |       |  | |  | || || |
Db         99 ATGCTTGCTTTCTTGCTTGCTTGCTTGCTTTCGTGCTTTCTTGCTTTCCCGTTTCTTGC 158

Qy        385 AGGGCTGCTCGC-TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 443
                  | ||  | |    | ||| |||||| |||  | ||||||||||||  |  || |
Db        159 TTTCTTTCTTTCTTTTGTTTCTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTT 218

Qy        444 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 503
              |||| || |||  |    |  ||  ||  |  ||  ||||||||||| || ||| || ||
Db        219 CTTGTTTTCTCCTTTCTTTCTTTCTTTTGTTTCTTTCCTGCTTGCTTTCTTGCTTGCTTG 278

Qy        504 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 542
              ||| | ||||  || || | |||| || ||  ||  | |
Db        279 CTTTCGTGCTTTCTTGCTTTCTTGTTTTCTTTCTTTCTT 317
```

RESULT 58

```
AAN50853
ID   AAN50853 standard; DNA; 600 BP.
XX
AC   AAN50853;
XX
DT   15-JAN-1992  (first entry)
XX
DE   Sequence encoding tetra-gastrin polymer.
XX
KW   E.coli; gastrin; ds.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   repeat_unit     1..12
FT                   /*tag= a
FT                   /number= 1
FT                   /note= "Repeated from 2-50 times"
XX
PN   JP60188081-A.
XX
PD   25-SEP-1985.
XX
PF   09-MAR-1984;   84JP-0046180.
XX
PR   09-MAR-1984;   84JP-0046180.
XX
PA   (SUMO ) SUMITOMO CHEM IND KK.
XX
DR   WPI; 1985-279157/45.
DR   P-PSDB; AAP50588.
XX
PT   Tetra-gastrin polymer gene - comprises many nucleotide sequence
PT   units of tetra-gastrin aminoacid sequence.
XX
PS   Claim 2; Page 491; 7pp; Japanese.
XX
CC   The repeat sequence encodes the tetra-gastrin polymer, and is useful
CC   in the large scale production of that protein, pref. through an E.
CC   coli host transformed with a pBR322 construct.
XX
SQ   Sequence 600 BP; 100 A; 100 C; 200 G; 200 T; 0 other;

  Query Match             6.9%;  Score 68.8;  DB 6;  Length 600;
  Best Local Similarity   44.7%;  Pred. No. 0.00024;
  Matches  268;  Conservative    0;  Mismatches  332;  Indels    0;  Gaps    0;

Qy        237 TGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 296
              |||   |  | |||   ||||    ||   |  | || | ||| || || | ||
Db          1 TGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTC 60

Qy        297 CAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAG 356
              |  || || || | ||        |  || || || | | |    |  | | |
Db         61 TGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTC 120

Qy        357 GGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGC 416
              |||  |       | |||  | || |  || || || | ||| || || | | |    |
Db        121 TGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTC 180

Qy        417 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 476
              |||   |  | || | | |    |  | | |  | || | ||| || || | | |    |
Db        181 TGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTC 240

Qy        477 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 536
              |||   |  | || | | |    |  | | |  | || | ||| || || | | |    |
Db        241 TGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTC 300

Qy        537 TGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGC 596
              |||   |  | |||   ||||    ||   |  | || | |       ||   | || |
Db        301 TGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTC 360

Qy        597 CTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 656
              |||||  |  | |  || || ||||   |      ||   |       | | |

Db        361 TGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTC 420

Qy        657 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGG 716
              | ||| || || |  || || ||||   |      ||        | | |
Db        421 TGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTC 480

Qy        717 CTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 776
              | ||| || || |  || || ||||   |      ||   |       | | |
Db        481 TGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTC 540

Qy        777 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGC 836
              | ||| || || |  || || ||||   |      ||   | || |||| | |  | |
Db        541 TGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTCTGGATGGACTTC 600


RESULT 59
AAS46422
ID   AAS46422 standard; DNA; 13321 BP.
XX
AC   AAS46422;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #144.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW   cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP02955.
XX
PR   15-MAR-2000; 2000DE-1013847.
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumour suppressor
PT   genes and oncogenes, useful in designing primers and probes for
PT   analysing diseases associated with cytosine methylation state e.g.
PT   cancer -
XX
PS   Claim 1; SEQ ID No 144; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and
CC   oncogenes having a sequence taken from 536 (actually 533 since
CC   numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC   (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC   peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC   form part of a set of probes for detecting the cytosine methylation state
CC   and/or single nucleotide polymorphisms and also to be used in an
CC   array for analysing diseases associated with CpG dinucleotides e.g.
CC   cancers and tumours. The probes can also be used in a method for
CC   ascertaining genetic and/or epigenetic parameters for the diagnosis
CC   and/or therapy of existing diseases or the predisposition to specific
CC   diseases, by analysing cytosine methylations. The parameters may be
CC   compared to another set of genetic and/or epigenetic parameters, the
CC   differences serving as basis for diagnosis and/or prognosis events which
CC   are disadvantageous to patients. The present sequence is one of the
```

```
CC   533 genomic sequences derived from tumour suppressor genes and
CC   oncogenes. Sequences with even numbered Seq ID numbers are the
CC   complementary sequence of the corresponding odd numbered sequence (e.g.
CC   ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC   is missing).
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 13321 BP; 2850 A; 580 C; 3835 G; 6056 T; 0 other;

  Query Match             6.9%;  Score 68.8;  DB 22;  Length 13321;
  Best Local Similarity   47.1%;  Pred. No. 0.00027;
  Matches  243;  Conservative    0;  Mismatches  272;  Indels    1;  Gaps    1;

Qy        249 TGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGC 308
              ||   |||     || | | | | | | |||   | | | | ||  ||  |  ||   |
Db      11247 TGGATGGATGAATGGACGGGTGGATGGATGGATGGATGGATAAGTAGATGGACGGGTGGG 11306

Qy        309 TTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGG 368
              | | | || |||  |  ||      | | | | |||   | | | ||   |   ||  |
Db      11307 TGGATAAGTGGATGGATGGATGAGTGGATGGATGGGTGGGTGGATGGATGAGTAGATAGA 11366

Qy        369 CTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 428
              | |||  | | | |  |||  | | |  ||   | |   | | | | |||   | | |
Db      11367 TGGGTGGGTGGATGGATGGGTGGGTGGACGGATGGATGAGTGGATAGATGGGTAGGTGGA 11426

Qy        429 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 488
              | | | | ||    | | | | | | | ||    | | | | | | | | |   | | |
Db      11427 TGGATGGATGAGTGGATGGATGGGTGGATGAATGGATGGATGGGTGGGTAGATGGATGGA 11486

Qy        489 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 548
              | | | |||  | | |   | | | | |||   | | | | | | | |||   | | |
Db      11487 TGGATGGATGGACGGGTGAGTAGATGGATGGGTGGGTGGATGGATGGATGGGTGGATGGA 11546

Qy        549 TGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGA 608
              |||| |||  | | || ||  ||  ||   ||  |   | | |  ||  |    | ||||
Db      11547 TGAGTGGA-TGGATGGGTGGGTGGATGGATAGATGGGTGAATGGATGGATAGATGATTGA 11605

Qy        609 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 668
              |||  || ||   || ||   | | ||  |   || ||   | | |   |   || ||
Db      11606 ATGTATGGGTGAGTGGATGGATGAATGGACGGATGGATGGATGGATGAGCGGATGGATGG 11665

Qy        669 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 728
              |   || ||   || ||    | | || |    || ||   | | || ||   || ||
Db      11666 GTGGGTGGATGGATGGATGAGTGGATGGATGGATGGATGGATGGATGAGCGGATGG 11725

Qy        729 TAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGG 764
              | | || |    || ||     | |  || | ||
Db      11726 ATGGATGGGTGGATGGATGGATAGATGTTTGTTCGG 11761


RESULT 60
AAX53491
ID   AAX53491 standard; DNA; 114955 BP.
XX
AC   AAX53491;
XX
DT   05-JUL-1999  (first entry)
XX
DE   Human adenosine A1 receptor antisense oligonucleotide fragment.
XX
KW   Antisense oligonucleotide; multiple target; antisense treatment;
KW   impaired respiration; inflammation; lung disease;
KW   pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW   acute asthma; allergy; asthma; impeded respiration;
KW   respiratory distress syndrome; pain; cystic fibrosis;
KW   pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW   chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW   colon cancer; breast cancer; lung cancer; pancreatic cancer;
KW   hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW   prostate cancer; ss.
XX
OS   Synthetic.
XX
PN   WO9913886-A1.
XX
PD   25-MAR-1999.
XX
PF   17-SEP-1998;   98WO-US19419.
XX
PR   09-JUN-1998;   98US-0093972.
PR   17-SEP-1997;   97US-0059160.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 1999-229400/19.
XX
PT   New antisense oligonucleotides used in treatment of, e.g. pulmonary
PT   vasoconstriction
XX
PS   Disclosure; Page 37; 120pp; English.
XX
CC   The specification describes antisense oligonucleotides (AAX52869-X55271)
CC   directed against at least 2 mRNAs selected from target genes, coding and
CC   non-coding regions of RNAs corresponding to target genes, gene
CC   initiation codons, genomic flanking regions, intron-exon borders, the
CC   5'-end, the 3'-end and the juxta-section between coding and non-coding
CC   regions and all segments of RNAs encoding proteins associated with one
CC   or more diseases, conditions or mixtures. The antisense oligonucleotides
CC   may be derived from sequences AAX55272-74. These multiple target
CC   oligonucleotides (specifically AAX55180-271) can be used for the
CC   antisense treatment of diseases and conditions. Typical diseases and
CC   conditions are those associated with impaired respiration and
CC   inflammation, including lung diseases, pulmonary vasoconstriction,
CC   inflammation, allergic rhinitis, acute asthma, allergies, asthma, impeded
CC   respiration, respiratory distress syndrome, pain, cystic fibrosis,
CC   pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic
CC   obstructive pulmonary disease (COPD), and cancers such as leukemias,
CC   lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer,
CC   pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma,
CC   hepatic metastases, as well as all types of cancers which may metastasize
CC   or have metastasized to the lungs, including breast and prostate cancer.
XX
SQ   Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;

  Query Match             6.7%;  Score 67.2;  DB 20;  Length 114955;
  Best Local Similarity   31.8%;  Pred. No. 0.00057;
  Matches  316;  Conservative   59;  Mismatches  619;  Indels    0;  Gaps    0;

Qy          3 CCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGT 62
              |||||||||  ||:       | |   |  |||  :       ||||||  |   ||
Db     107887 CCGCCCGGCCCGGSNNNDNNGCCCGCCCGGCCCGGSNNNDNNCCCGCCCGGCCCGGSNNN 107946

Qy         63 CTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGG 122
                 |  |||  |  | :   :  ||     |  ||:   : ||  |||||||:   :
Db     107947 DNNCCGCCCGGCCCGGSNNNDNNCGCCCGGCCCGGSNNNDNNGCCCGGCCCGGSNNNDNN 108006

Qy        123 CGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAG 182
              |||   | ||   |     |  ||   :           :   | || |  || | |
Db     108007 GGCGGGGGCGGCGGCGCCTGGCTCGCCTBGGGCCCCSNNNDNNGGCGGGGGCGGCGGCGC 108066

Qy        183 GGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCT 242
                | |      :| |  |    :       |||    |  ||||| ||   | : ||||
Db     108067 CTGGCTCGCCTBGGGCCCSNNNDNNGGCGGGGGCGGCGGCGCCTGGCTCGCCTBGGGCCS 108126

Qy        243 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAG 302
                :   | |||| | |  |  |       | |  :|| |    :   | |||| | |  |
Db     108127 NNNDNNGGCGGGGGCGGCGGCGCCTGGCTCGCCTBGGGCSNNNDNNGGCGGGGGCGGCGG 108186

Qy        303 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 362
              |     |    :|| :   :    |   ||      |||||  |     :||:
Db     108187 CGCCTGGCTCGCCTBGGGSNNNDNNGGCGGGGGCGGCGGCGCCTGGCTCGCCTBGGSNNN 108246
```

```
Qy        363 GTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 422
              :  ||   |   |   ||   | |  |  ||:|:      |   |      ||  | |
Db     108247 DNNGGCGGGGGCGGCGGCGCCTGGCTCGCCTBGSNNNDNNGGCGGGGGCGGCGGCGCCTG 108306

Qy        423 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 482
              ||| || |::      | | |       |    |||||  || |:   :   || ||
Db     108307 GCTCGCCTBSNNNDNNGGCGGGGGCGGCGGCGCCTGGCTCGCCTSNNNDNNGGCGGGGGC 108366

Qy        483 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 542
              |   ||         | |:   :     |   || |    || ||  |   :   :
Db     108367 GGCGGCGCCTGGCTCGCCSNNNDNNGGCGGGGGCGGCGGCGCCTGGCTCGCSNNNDNNGG 108426

Qy        543 GCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATG 602
              |  |  ||  |            :          | ||  |   |     |:   :
Db     108427 CGGGGGCGGCGGCGCCTGGCTCGSNNNDNNGGCGGGGGCGGCGGCGCCTGGCTCSNNNDN 108486

Qy        603 GCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 662
              |   | | |  ||  ||||:   :  ||   |   ||| ||   | | :
Db     108487 NGGCGGGGGCGGCGGCGCCTGGCTSNNNDNNGGCGGGGGCGGCGGCGCCTGGCSNNNDNN 108546

Qy        663 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTG 722
              ||| |       |||  ||||:       |  |||  |   |||  || :   :   || |
Db     108547 GGCGGGGGCGGCGGCGCCTGGSNNNDNNGGCGGGGGCGGCGGCGCCTGSNNNDNNGGCGG 108606

Qy        723 GGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 782
              ||  |   |     : . :  |  |   |  | |  || :   :       ||||  || |
Db     108607 GGGCGGCGGCGCCTSNNNDNNGGCGGGGGCGGCGGCGCCSNNNDNNGGCGGGGGCGGCGG 108666

Qy        783 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTT 842
               ||    :   | |  |  ||         :  | .|   || |    :   :   ||
Db     108667 CGCSNNNDNNGGCGGGGGCGGCGGCGSNNNDNNGGCGGGGGCGGCGGCSNNNDNNGGCGG 108726

Qy        843 GCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGGG 902
              |   |  ||:       | | |   |   |:       || |  | |       :     ||
Db     108727 GGGCGGCGGSNNNDNNGGCGGGGGCGGCGSNNNDNNGGCGGGGGCGGCSNNNDNNGGCGG 108786

Qy        903 CTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACCCGGGCC 962
                || |:   :  ||| | | ||  |  |  ||  |    : || | |:       ||
Db     108787 GGGCGGSNNNDNNGCGGGGGCGGCGGCGCCTGGCTCGCCTBGGGCCCCSNNNDNNCGGGG 108846

Qy        963 TCTCCCCCGCCCGGACCGCCGGGCACCCGGCCCA 996
              |  | |||| ||  |||| :|  ||| :   :
Db     108847 GCGGCGGCGCCTGGCTCGCCTBGGGCCCCSNNND 108880
```

RESULT 61
ABQ97197
ID ABQ97197 standard; cDNA; 397 BP.
XX
AC ABQ97197;
XX
DT 30-OCT-2002 (first entry)
XX
DE Mouse ES cell related cDNA SEQ ID NO 465.
XX
KW Mouse; ES cell; gene trapped sequence; GTS; gene expression;
KW development disorder; cell differentiation disorder; gene; ss.
XX
OS Mus sp.
XX
PN US2002081668-A1.
XX
PD 27-JUN-2002.
XX
PF 30-NOV-2000; 2000US-0728446.
XX
PR 20-NOV-1998; 98US-109302P.
PR 01-DEC-1999; 99US-168270P.
XX
PA (FRIE/) FRIEDRICH G.
PA (ZAMB/) ZAMBROWICZ B.
PA (SAND/) SANDS A T.
XX
PI Friedrich G, Zambrowicz B, Sands AT;
XX
DR WPI; 2002-626541/67.
XX
PT Novel murine polynucleotides that individually identify novel genes
PT into which a retroviral gene trap vector has been integrated, useful in
PT genomic analysis and in discovery, development of therapeutic and
PT diagnostic agents -
XX
PS Claim 2; SEQ ID NO 465; 29pp + Sequence Listing; English.
XX
CC The invention relates to isolated murine polynucleotides (I) comprising a
CC contiguous stretch of at least about 60 nucleotides of a sequence
CC (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences
CC (GTSs). The novel genes can be used in a process to identify novel
CC polynucleotide sequences by comparing them to the novel gene sequences.
CC The novel genes and cells are useful in functional genomic analysis and
CC in the discovery and development of new therapeutic and diagnostic agents
CC and methods. (I) is useful for identifying the coding regions of the
CC murine genome, to isolate cDNAs, genomic clones or full-length
CC genes/polynucleotides or homologues, heterologues, paralogues or
CC orthologues that are capable of hybridising to one or more of the GTSs
CC under stringent conditions. (I) can be incorporated into a phage display
CC system that can be used to screen for proteins or other ligands, that are
CC capable of binding an amino acid sequence encoded by an oligonucleotide
CC or polynucleotide sequence in at least one of the GTS sequences. (I) is
CC useful in arrays, such as gene chips, to identify and characterise
CC temporal and tissue specific gene expression, to identify the gene of
CC interest from many sources and for genetic manipulations such as
CC antisense inhibition and gene targeting. Decreasing the level of
CC expression of (I) and/or down regulating the activity of peptides or
CC proteins encoded by (I) is useful for treating development and cell
CC differentiation disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at sequdata.uspto.gov/sequence.html?DocID=20020081668.
XX
SQ Sequence 397 BP; 89 A; 76 C; 115 G; 113 T; 4 other;

  Query Match   6.7%; Score 67; DB 24; Length 397;
  Best Local Similarity   65.5%; Pred. No. 0.00053;
  Matches   97; Conservative   0; Mismatches   51; Indels   0; Gaps   0;

```
Qy        392 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 451
              ||| |||   | | || |  | |  |||  |||| | ||||| | | ||  | |||||
Db         32 CTCCCTGTTGTTCCTGTTGCCACTTTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGG 91

Qy        452 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 511
              | |||||  |||| | ||||| | | ||  | ||||| | | |||||  |||| | |||
Db         92 ATTGCTGGATTGCTGGATTGCTGGATTGCTGGATNGCTGGGTTGCTGGGTTG 151

Qy        512 CTCGCTGGCCTGCTTGCTTGCTCGCTGG 539
              || | | ||  | ||||| | | |||||
Db        152 CTGGATTGCTGGATTGCTGGATTGCTGG 179
```

RESULT 62
AAL40271/c
ID AAL40271 standard; DNA; 183 BP.
XX
AC AAL40271;
XX
DT 13-SEP-2002 (first entry)
XX
DE BASB204 related DNA sequence SEQ ID No 31.
XX
KW Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW hearing loss; chronic otitis media; fluid accumulation; inflammation;

```
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   P-PSDB; AAO21918.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear  -
XX
PS   Example 2; Fig 3; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 183 BP; 69 A; 36 C; 33 G; 45 T; 0 other;

  Query Match             6.7%;  Score 66.8;  DB 24;  Length 183;
  Best Local Similarity   86.0%;  Pred. No. 0.00056;
  Matches   74;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        401 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 460
              ||||||||||||| ||| || ||||||||||||| ||| || ||||||||||||| ||| ||
Db         96 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 37

Qy        461 CTGCTTGCTTGCTCGCTGGCCTGCTT 486
              ||||||||||||| ||| || |||||
Db         36 TTGCTTGCTTGCTTGCTTGCTTGCTT 11


RESULT 63
AAL40264/c
ID   AAL40264 standard; DNA; 186 BP.
XX
AC   AAL40264;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 17.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   P-PSDB; AAO21911.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear  -
XX
PS   Example 2; Fig 3; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 186 BP; 71 A; 36 C; 33 G; 46 T; 0 other;

  Query Match             6.7%;  Score 66.8;  DB 24;  Length 186;
  Best Local Similarity   86.0%;  Pred. No. 0.00056;
  Matches   74;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        401 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 460
              ||||||||||||| ||| || ||||||||||||| ||| || ||||||||||||| ||| ||
Db         96 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 37

Qy        461 CTGCTTGCTTGCTCGCTGGCCTGCTT 486
              ||||||||||||| ||| || |||||
Db         36 TTGCTTGCTTGCTTGCTTGCTTGCTT 11


RESULT 64
AAL40265/c
```

ID AAL40265 standard; DNA; 228 BP.
XX
AC AAL40265;
XX
DT 13-SEP-2002 (first entry)
XX
DE BASB204 related DNA sequence SEQ ID No 19.
XX
KW Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW auditive nerve damage; delayed speech learning; gene; ds.
XX
OS Haemophilus influenzae.
XX
PN WO200234768-A2.
XX
PD 02-MAY-2002.
XX
PF 24-OCT-2001; 2001WO-EP12390.
XX
PR 24-OCT-2000; 2000GB-0026002.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Thonnard J;
XX
DR WPI; 2002-471431/50.
DR P-PSDB; AAO21912.
XX
PT New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT useful as components of vaccines useful for treating bacterial
PT infection such as otitis media, delayed speech learning, and
PT inflammation of the middle ear -
XX
PS Example 2; Fig 3; 97pp; English.
XX
CC The invention relates to an isolated polypeptide comprising an amino acid
CC sequence which has 85% identity to the entire length of a fully defined
CC non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC in specification. The isolated protein, an immunogenic fragment of the
CC protein or an antibody immunospecific for the protein or immunogenic
CC fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC The molecules may be used in the prevention of adhesion of Gram-negative
CC or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC wounds, to block bacterial adhesion between mammalian extracellular
CC matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC block the normal progression of pathogenesis in infections initiated
CC other than by the implantation of in-dwelling devices. The isolated
CC protein encoding polynucleotide is useful as a research reagent and
CC material for discovery of treatments of and diagnostics for diseases,
CC particularly human diseases. The antibody of the invention is useful for
CC treating infections preferably bacterial infections such as otitis media
CC in infants and children, pneumonia in elders, sinusitis, nosocomial
CC infections, or invasive diseases, chronic otitis media with hearing loss,
CC fluid accumulation in the middle ear, auditive nerve damage, delayed
CC speech learning, infections of the upper respiratory tract and
CC inflammation of the middle ear. This polynucleotide sequence represents a
CC Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ Sequence 228 BP; 93 A; 42 C; 37 G; 56 T; 0 other;

Query Match 6.7%; Score 66.8; DB 24; Length 228;
Best Local Similarity 86.0%; Pred. No. 0.00056;
Matches 74; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 401 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 460
||||||||||||| ||| || ||||||||||||| ||| || ||||||||||||| ||| ||
Db 96 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 37

Qy 461 CTGCTTGCTTGCTCGCTGGCCTGCTT 486
||||||||||||| ||| || |||||
Db 36 TTGCTTGCTTGCTTGCTTGCTTGCTT 11

RESULT 65
AAL40268/c
ID AAL40268 standard; DNA; 240 BP.
XX
AC AAL40268;
XX
DT 13-SEP-2002 (first entry)
XX
DE BASB204 related DNA sequence SEQ ID No 25.
XX
KW Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW auditive nerve damage; delayed speech learning; gene; ds.
XX
OS Haemophilus influenzae.
XX
PN WO200234768-A2.
XX
PD 02-MAY-2002.
XX
PF 24-OCT-2001; 2001WO-EP12390.
XX
PR 24-OCT-2000; 2000GB-0026002.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Thonnard J;
XX
DR WPI; 2002-471431/50.
DR P-PSDB; AAO21915.
XX
PT New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT useful as components of vaccines useful for treating bacterial
PT infection such as otitis media, delayed speech learning, and
PT inflammation of the middle ear -
XX
PS Example 2; Fig 3; 97pp; English.
XX
CC The invention relates to an isolated polypeptide comprising an amino acid
CC sequence which has 85% identity to the entire length of a fully defined
CC non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC in specification. The isolated protein, an immunogenic fragment of the
CC protein or an antibody immunospecific for the protein or immunogenic
CC fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC The molecules may be used in the prevention of adhesion of Gram-negative
CC or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC wounds, to block bacterial adhesion between mammalian extracellular
CC matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC block the normal progression of pathogenesis in infections initiated
CC other than by the implantation of in-dwelling devices. The isolated
CC protein encoding polynucleotide is useful as a research reagent and
CC material for discovery of treatments of and diagnostics for diseases,
CC particularly human diseases. The antibody of the invention is useful for
CC treating infections preferably bacterial infections such as otitis media
CC in infants and children, pneumonia in elders, sinusitis, nosocomial
CC infections, or invasive diseases, chronic otitis media with hearing loss,
CC fluid accumulation in the middle ear, auditive nerve damage, delayed
CC speech learning, infections of the upper respiratory tract and
CC inflammation of the middle ear. This polynucleotide sequence represents a
CC Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ Sequence 240 BP; 98 A; 44 C; 40 G; 58 T; 0 other;

Query Match 6.7%; Score 66.8; DB 24; Length 240;

```
  Best Local Similarity   86.0%;  Pred. No. 0.00057;
  Matches   74;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        401 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 460
              ||||||||||||| ||| || |||||||||||| ||| || |||||||||||| ||| ||
Db         96 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 37

Qy        461 CTGCTTGCTTGCTCGCTGGCCTGCTT 486
              ||||||||||||| ||| || |||||
Db         36 TTGCTTGCTTGCTTGCTTGCTTGCTT 11


RESULT 66
AAL40266/c
ID   AAL40266 standard; DNA; 252 BP.
XX
AC   AAL40266;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 21.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   P-PSDB; AAO21913.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear  -
XX
PS   Example 1; Fig 3; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 252 BP; 101 A; 45 C; 45 G; 61 T; 0 other;

  Query Match             6.7%;  Score 66.8;  DB 24;  Length 252;
  Best Local Similarity   86.0%;  Pred. No. 0.00057;
  Matches   74;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        401 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 460
              ||||||||||||| ||| || |||||||||||| ||| || |||||||||||| ||| ||
Db         96 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 37

Qy        461 CTGCTTGCTTGCTCGCTGGCCTGCTT 486
              ||||||||||||| ||| || |||||
Db         36 TTGCTTGCTTGCTTGCTTGCTTGCTT 11


RESULT 67
AAL40267/c
ID   AAL40267 standard; DNA; 264 BP.
XX
AC   AAL40267;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 23.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   P-PSDB; AAO21914.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear  -
XX
PS   Example 2; Fig 3; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
```

CC protein encoding polynucleotide is useful as a research reagent and
CC material for discovery of treatments of and diagnostics for diseases,
CC particularly human diseases. The antibody of the invention is useful for
CC treating infections preferably bacterial infections such as otitis media
CC in infants and children, pneumonia in elders, sinusitis, nosocomial
CC infections, or invasive diseases, chronic otitis media with hearing loss,
CC fluid accumulation in the middle ear, auditive nerve damage, delayed
CC speech learning, infections of the upper respiratory tract and
CC inflammation of the middle ear. This polynucleotide sequence represents a
CC Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ Sequence 264 BP; 110 A; 45 C; 46 G; 63 T; 0 other;

Query Match 6.7%; Score 66.8; DB 24; Length 264;
Best Local Similarity 86.0%; Pred. No. 0.00057;
Matches 74; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 401 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 460
||||||||||||| ||| || ||||||||||||| ||| || ||||||||||||| ||| ||
Db 96 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 37

Qy 461 CTGCTTGCTTGCTCGCTGGCCTGCTT 486
||||||||||||| ||| || |||||
Db 36 TTGCTTGCTTGCTTGCTTGCTTGCTT 11

RESULT 68
AAL40269/c
ID AAL40269 standard; DNA; 267 BP.
XX
AC AAL40269;
XX
DT 13-SEP-2002 (first entry)
XX
DE BASB204 related DNA sequence SEQ ID No 27.
XX
KW Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW auditive nerve damage; delayed speech learning; gene; ds.
XX
OS Haemophilus influenzae.
XX
PN WO200234768-A2.
XX
PD 02-MAY-2002.
XX
PF 24-OCT-2001; 2001WO-EP12390.
XX
PR 24-OCT-2000; 2000GB-0026002.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Thonnard J;
XX
DR WPI; 2002-471431/50.
DR P-PSDB; AAO21916.
XX
PT New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT useful as components of vaccines useful for treating bacterial
PT infection such as otitis media, delayed speech learning, and
PT inflammation of the middle ear -
XX
PS Example 2; Fig 3; 97pp; English.
XX
CC The invention relates to an isolated polypeptide comprising an amino acid
CC sequence which has 85% identity to the entire length of a fully defined
CC non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC in specification. The isolated protein, an immunogenic fragment of the
CC protein or an antibody immunospecific for the protein or immunogenic
CC fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC The molecules may be used in the prevention of adhesion of Gram-negative
CC or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC wounds, to block bacterial adhesion between mammalian extracellular
CC matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC block the normal progression of pathogenesis in infections initiated
CC other than by the implantation of in-dwelling devices. The isolated
CC protein encoding polynucleotide is useful as a research reagent and
CC material for discovery of treatments of and diagnostics for diseases,
CC particularly human diseases. The antibody of the invention is useful for
CC treating infections preferably bacterial infections such as otitis media
CC in infants and children, pneumonia in elders, sinusitis, nosocomial
CC infections, or invasive diseases, chronic otitis media with hearing loss,
CC fluid accumulation in the middle ear, auditive nerve damage, delayed
CC speech learning, infections of the upper respiratory tract and
CC inflammation of the middle ear. This polynucleotide sequence represents a
CC Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ Sequence 267 BP; 111 A; 45 C; 46 G; 65 T; 0 other;

Query Match 6.7%; Score 66.8; DB 24; Length 267;
Best Local Similarity 86.0%; Pred. No. 0.00057;
Matches 74; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 401 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 460
||||||||||||| ||| || ||||||||||||| ||| || ||||||||||||| ||| ||
Db 96 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 37

Qy 461 CTGCTTGCTTGCTCGCTGGCCTGCTT 486
||||||||||||| ||| || |||||
Db 36 TTGCTTGCTTGCTTGCTTGCTTGCTT 11

RESULT 69
AAL40270/c
ID AAL40270 standard; DNA; 267 BP.
XX
AC AAL40270;
XX
DT 13-SEP-2002 (first entry)
XX
DE BASB204 related DNA sequence SEQ ID No 29.
XX
KW Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW auditive nerve damage; delayed speech learning; gene; ds.
XX
OS Haemophilus influenzae.
XX
PN WO200234768-A2.
XX
PD 02-MAY-2002.
XX
PF 24-OCT-2001; 2001WO-EP12390.
XX
PR 24-OCT-2000; 2000GB-0026002.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Thonnard J;
XX
DR WPI; 2002-471431/50.
DR P-PSDB; AAO21917.
XX
PT New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT useful as components of vaccines useful for treating bacterial
PT infection such as otitis media, delayed speech learning, and
PT inflammation of the middle ear -
XX

PS Example 2; Fig 3; 97pp; English.
XX
CC The invention relates to an isolated polypeptide comprising an amino acid
CC sequence which has 85% identity to the entire length of a fully defined
CC non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC in specification. The isolated protein, an immunogenic fragment of the
CC protein or an antibody immunospecific for the protein or immunogenic
CC fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC The molecules may be used in the prevention of adhesion of Gram-negative
CC or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC wounds, to block bacterial adhesion between mammalian extracellular
CC matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC block the normal progression of pathogenesis in infections initiated
CC other than by the implantation of in-dwelling devices. The isolated
CC protein encoding polynucleotide is useful as a research reagent and
CC material for discovery of treatments of and diagnostics for diseases,
CC particularly human diseases. The antibody of the invention is useful for
CC treating infections preferably bacterial infections such as otitis media
CC in infants and children, pneumonia in elders, sinusitis, nosocomial
CC infections, or invasive diseases, chronic otitis media with hearing loss,
CC fluid accumulation in the middle ear, auditive nerve damage, delayed
CC speech learning, infections of the upper respiratory tract and
CC inflammation of the middle ear. This polynucleotide sequence represents a
CC Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ Sequence 267 BP; 111 A; 45 C; 46 G; 65 T; 0 other;

Query Match 6.7%; Score 66.8; DB 24; Length 267;
Best Local Similarity 86.0%; Pred. No. 0.00057;
Matches 74; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 401 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 460
||||||||||||| ||| || ||||||||||||| ||| || ||||||||||||| ||| ||
Db 96 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 37

Qy 461 CTGCTTGCTTGCTCGCTGGCCTGCTT 486
||||||||||||| ||| || |||||
Db 36 TTGCTTGCTTGCTTGCTTGCTT 11

RESULT 70
ABA17155/c
ID ABA17155 standard; DNA; 32249 BP.
XX
AC ABA17155;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 9486.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 9486; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 32249 BP; 7365 A; 9042 C; 8692 G; 7150 T; 0 other;

Query Match 6.7%; Score 66.6; DB 22; Length 32249;
Best Local Similarity 45.0%; Pred. No. 0.00071;
Matches 328; Conservative 0; Mismatches 399; Indels 2; Gaps 2;

```
Qy       145 ACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGT 204
Db     13050 ATGGATGGGTGGGAGGAGGCAGTGAGGGAGGGAAGGAAGGAGGCATGGATGAATGCATGG 12991

Qy       205 GAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGC 264
Db     12990 AGGAAAGAATGAATGAATGAATGAATGAATGAATGAATGAATGCATGGGTGAATGGATGG 12931

Qy       265 CTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGT 324
Db     12930 ATGGGTGCATGGATGCATGGAGGAATGGATGGATGGATGGATGCATGCATGGAGAAATGG 12871

Qy       325 AGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAG-GGGACAGTAGGGCTGCTGGCTTGCTAG 383
Db     12870 ATGGATGGGTGCATGGAGGAATGGATGGAGGGATGCACGGAAGAATGGGCAGATGGATGG 12811

Qy       384 TAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 443
Db     12810 ATGGATGGATGGAAGAATCAATGGAGGGAGTGATGGATAGATGGATTGATGGATGGATGG 12751

Qy       444 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 503
Db     12750 ATGGATGGATGGATGGACGAGTGGATGGATGGGTAAATGGATGGATGGATGAGTGGATGG 12691

Qy       504 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGG 563
Db     12690 ATGAATGGATGGATGG-ATAGATGGATGCATGGAGGAATGGATGCATGGAGGAATGGATG 12632

Qy       564 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 623
Db     12631 GATGAGTGGATGGATAGAGGGAAGAATGGATGAATGAGTGGATGGATAGAGGGAAGAATG 12572

Qy       624 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 683
Db     12571 GATGGAAGGAAGAATTGATGGGTAGATGGATGGATGGATGGATAGATGGGTGAGTGGATG 12512

Qy       684 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 743
Db     12511 GATGGATGGATGGATGGATGGATGGATGGAAGGATGCATGGAGAAATGGATGGATGGGTG 12452

Qy       744 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 803
Db     12451 GATGGATGCATGGAGGAATAGATGGATGGATGGATGGATGGATGAATGGATGGAAGGATG 12392

Qy       804 GCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTG 863
Db     12391 GATGCATGCATGGAGAAATGGATGGGTGGGTGCATGGAGGAATGGATGGAGGGATGCATG 12332

Qy       864 CTACTAGGG 872
Db     12331 GAAGAATGG 12323
```

```
RESULT 71
ABZ40780/c
ID   ABZ40780 standard; DNA; 147 BP.
XX
AC   ABZ40780;
XX
DT   07-MAR-2003  (first entry)
XX
DE   N. gonorrhoeae nucleotide sequence SEQ ID 6149.
XX
KW   Antibacterial; infection; vaccine; gene therapy; gene; ds.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO200279243-A2.
XX
PD   10-OCT-2002.
XX
PF   12-FEB-2002; 2002WO-IB02069.
XX
PR   12-FEB-2001; 2001GB-0003424.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Fontana MR,  Pizza M,  Masignani V,  Monaci E;
XX
DR   WPI; 2003-058415/05.
DR   P-PSDB; ABP79810.
XX
PT   New protein from Neisseria gonorrheae, useful for the manufacture of a
PT   medicament for treating or preventing N. gonorrheae infection  -
XX
PS   Disclosure; Page 637; 815pp; English.
XX
CC   The present invention relates to proteins from Neisseria gonorrhoeae.
CC   Also disclosed are the nucleic acid molecules encoding the proteins and
CC   antibodies that specifically bind to the proteins. The composition
CC   comprising the protein, nucleic acid or antibody is useful for the
CC   manufacture of a medicament for treating or preventing N. gonorrhoeae
CC   infection, this may be in the form of a vaccine or gene therapy.
CC   Sequences given in records ABZ37706-ABZ42016 represent nucleic acid
CC   molecules of the invention.
XX
SQ   Sequence 147 BP; 66 A; 35 C; 31 G; 15 T; 0 other;

  Query Match             6.6%;  Score 66;  DB 25;  Length 147;
  Best Local Similarity   79.6%;  Pred. No. 0.00079;
  Matches   78;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy        397 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 456
              | || |||||||||||| ||| || |||||||||||| ||| || |||||||||||| ||
Db        145 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 86

Qy        457 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 494
              | || |||||||||||| ||| | ||| ||   |  ||
Db         85 TTGCTTGCTTGCTTGCTTGCTTGACTGTTTTTGTTTTC 48


RESULT 72
AAS76504
ID   AAS76504 standard; cDNA; 3249 BP.
XX
AC   AAS76504;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #12308.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG12317.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 12308; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3249 BP; 631 A; 1058 C; 972 G; 588 T; 0 other;

  Query Match             6.5%;  Score 65.4;  DB 23;  Length 3249;
  Best Local Similarity   49.0%;  Pred. No. 0.0011;
  Matches  174;  Conservative    0;  Mismatches  181;  Indels    0;  Gaps    0;

Qy        375 GCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 434
              || ||   |  ||  || || |  |||||| || | ||  | |  || || ||  || |
Db       2861 GCATGTCTGATGGCGTGGTCCCCAGCCTGCCTGATGGCGTGGTCCCCAGCCTGTCTGATG 2920

Qy        435 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 494
              ||  |    |  || ||   | |||| || |  |  |   |   |   || || |  |
Db       2921 GCATGGTCCCCGGCCTGTCTGATGGCGTGGTCCCCAGTCTGTCTGATGGCGTGGTCCCCG 2980

Qy        495 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 554
              ||  | ||| | || || || | ||||||  || | ||  | |  || | ||      |  .
Db       2981 GCCTGTCTGATGGCGTGGTCCCCGGCCTGTCTGATGGCATGGTCCCCGGTCTGTCTGATG 3040

Qy        555 GACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAG 614
              |   |     | || ||  || |  | |     ||||| |||||||||| |    |
Db       3041 GTGTGGTCCCCAGCCTGTCTGATGGTGTGGTCCCCGGCCTGTCTGATGGCTTGGTCCCTG 3100

Qy        615 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 674
              | ||  | |   |   |     | ||  | |||  ||  ||   |  | |||   |  |
Db       3101 GTCTATTTGATGGTGTGGTCCCCAGCCTGTCTGAAGGTGTGGTCCCCAGCCTGTCTGATG 3160

Qy        675 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 729
```

```
Db       3161 GCGTGGTCCCCGGCCTGTCTGATGGCTTGGTCCCTGGTCTATTTGATGGTGTGGT 3215


RESULT 73
AAV88624/c
ID   AAV88624 standard; cDNA; 368 BP.
XX
AC   AAV88624;
XX
DT   12-FEB-1999  (first entry)
XX
DE   EST clone GX576.
XX
KW   Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW   tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW   chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW   receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS   Homo sapiens.
XX
PN   WO9845437-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06956.
XX
PR   10-APR-1997;   97US-0837312.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070078/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries
XX
PS   Claim 1; Page 458; 641pp; English.
XX
CC   The present sequence represents an expressed sequence tag (EST), and is
CC   a polynucleotide of the invention. The polynucleotides of the invention
CC   are all secreted EST sequences isolated from a variety of human tissue
CC   sources. The EST sequences and proteins encoded by them are predicted to
CC   have useful biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals, although no supporting data is given. Suggested activities
CC   include nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The EST sequences are also stated to be useful for gene
CC   therapy.
XX
SQ   Sequence 368 BP; 216 A; 51 C; 89 G; 9 T; 3 other;

  Query Match             6.5%;  Score 65.2;  DB 20;  Length 368;
  Best Local Similarity   68.5%;  Pred. No. 0.0011;
  Matches  102;  Conservative    0;  Mismatches   46;  Indels    1;  Gaps    1;

Qy        402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461
Db        231 TTCTTTCTTGCTTGCTTTCTTGCTTGCTNGCTTGCTTTCGTGCTTTCTTGCTTTCCTGTT 172

Qy        462 TGCTTGCTTGCTCGCTGGCCTGC-TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 520
Db        171 TNCTTTCTTTCTTTCTTTCTTTTGTTTCTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTC 112

Qy        521 CTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
Db        111 GTGCTTTCTTGTTTTCTCGCNTTCTTTCT 83


RESULT 74
AAT62138/c
ID   AAT62138 standard; DNA; 1771 BP.
XX
AC   AAT62138;
XX
DT   13-JUL-1997  (first entry)
XX
DE   Leishmania tropica Lt-210 antigen cDNA.
XX
KW   Leishmaniasis; vaccine; diagnosis; Lt-210; antigen; immunogen; ss.
XX
OS   Leishmania tropica strain MHOM/SA/91/WR1063C.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1701
FT                   /*tag= a
XX
PN   WO9711180-A1.
XX
PD   27-MAR-1997.
XX
PF   20-SEP-1996;   96WO-US15185.
XX
PR   22-SEP-1995;   95US-0533669.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Campos-Neto A,  Dillon DC,  Reed SG,  Skeiky YA,  Webb JR;
XX
DR   WPI; 1997-202888/18.
DR   P-PSDB; AAW11217.
XX
PT   Polypeptide(s) comprising immunogenic part of Leishmania antigen -
PT   useful for leishmaniasis vaccination and diagnosis
XX
PS   Example 4; Page 64-67; 106pp; English.
XX
CC   A genomic DNA molecule (AAT62138) encodes a 210 kDa immunogenic
CC   Leishmania tropica antigen (AAW11217), designated Lt-210.  It was
CC   obtd. by screening a L. tropica promastigote expression library
CC   with sera isolated from viscerotropic leishmaniasis patients.  DNA
CC   molecules (see also AAT62135-37, AAT62139) encoding Leishmania antigens
CC   can be incorporated into vectors and used to produce recombinant
CC   antigens, or immunogenic portions of them, in host cells for use in
CC   vaccines against Leishmaniasis, or may themselves be used as
CC   vaccines for in vivo immunogen generation.
XX
SQ   Sequence 1771 BP; 279 A; 552 C; 672 G; 268 T; 0 other;

  Query Match             6.4%;  Score 64.2;  DB 18;  Length 1771;
  Best Local Similarity   48.7%;  Pred. No. 0.0018;
  Matches  320;  Conservative    0;  Mismatches  328;  Indels    9;  Gaps    5;

Qy        207 GACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 266
Db        881 GCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGTGCATGACATCCAGGCGGGCC 822

Qy        267 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 326
Db        821 GCCTCCTCGGCCGCCT-CCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGAC 763

Qy        327 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 386
Db        762 GCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAG 703

Qy        387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 446
Db        702 CTCCGCGCGGAGACGCGC-TGCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGT 644
```

```
Qy        447 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 506
              || ||    |  | |     || | ||||  ||||||    |  ||  ||| | |  | |
Db        643 GCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGGCCTCGT 584

Qy        507 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT---GCTTGCTGAGGGGACAGTAGG 563
              || ||    |  | | |   || | ||||   ||||   |||    | || |||     |
Db        583 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTG 524

Qy        564 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 623
              || || |   ||  || |  ||   ||  | |     || ||       ||| |   |
Db        523 CCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGCATGGCCTCCAGGCGGGCCG 464

Qy        624 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 683
              || || |     || | |  | ||  | ||    ||| | || || |  || |||   |
Db        463 CCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-GCCTCCTCGAGGGCCTGGACG 405

Qy        684 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 743
              || |   |||  | |  ||  |    ||| |   | || || |     |||| |
Db        404 CGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCGCCTCCTCGGCTTCCTCCAGC 345

Qy        744 GCTGC---TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 800
              | ||    ||  | |||||     |   |  ||  || || ||||| |     |||
Db        344 TCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGC 285

Qy        801 TGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 857
              |||  | ||  | || |||| || |   || | |      ||  |  |   ||||
Db        284 ATGGCCTCCAGGCGCGCCGCCTCCTCGGCAGCCTCCAGCTCCGCGCGGAGACGCGCT 228
```

RESULT 75
AAV47558/c
ID AAV47558 standard; DNA; 1771 BP.
XX
AC AAV47558;
XX
DT 13-NOV-1998 (first entry)
XX
DE Leishmania antigen Lt-210 coding sequence.
XX
KW Leishmania antigen; immune response; infection detection; therapy;
KW humoral response induction; cellular response induction; cancer;
KW interleukin-12 production; ss.
XX
OS Leishmania sp.
XX
FH Key Location/Qualifiers
FT CDS 1..1701
FT /*tag= a
FT /product= Lt-210
XX
PN WO9835045-A2.
XX
PD 13-AUG-1998.
XX
PF 12-FEB-1998; 98WO-US03002.
XX
PR 27-AUG-1997; 97US-0920609.
PR 12-FEB-1997; 97US-0798841.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Reed SG, Skeiky YA, Webb JR;
XX
DR WPI; 1998-447242/38.
DR P-PSDB; AAW70212.
XX
PT New immunogenic fragments of Leishmania antigens and related nucleic
PT acid, vectors and host cells - are useful for diagnosis, prevention
PT and treatment of leishmaniasis, also to induce production of
PT interleukin-12 generally
XX
PS Disclosure; Page 79-82; 194pp; English.
XX
CC This sequence encodes a Leishmania antigen (LAg) of the invention,
CC designated Lt-210. Compositions and vaccines containing the protein are
CC used to generate a protective or therapeutic immune response against the
CC Leishmania species donavani, chagasi, infantum, major, amazonensis,
CC braziliensis, panamensis, tropica or guayanensis. They can also be used
CC to detect infection (in a skin test). The compositions induce a humoral
CC and/or cellular response, specifically of Th1 type, particularly
CC including induction of interleukin-12 (IL-12) production. They may thus
CC be used more generally to treat any condition (e.g. bacterial, viral or
CC protozoal infection, or cancer) which responds to IL-12.
XX
SQ Sequence 1771 BP; 279 A; 552 C; 672 G; 268 T; 0 other;

Query Match 6.4%; Score 64.2; DB 19; Length 1771;
Best Local Similarity 48.7%; Pred. No. 0.0018;
Matches 320; Conservative 0; Mismatches 328; Indels 9; Gaps 5;

```
Qy        207 GACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 266
              | | ||  |   | |||    |  ||  ||| |||  | |||         |||  || |
Db        881 GCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGTGCATGACATCCAGGCGGGCC 822

Qy        267 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 326
              || | ||||   |||| |  |||  | |   ||  |  |||| | ||   ||| | | |
Db        821 GCCTCCTCGGCCGCCT-CCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGAC 763

Qy        327 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 386
              ||  || ||||||    | ||| ||       | | |   | || || |   ||    ||
Db        762 GCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAG 703

Qy        387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 446
                | ||  ||  | |  || ||| | ||||  ||||||   ||  ||  ||| | |  | |
Db        702 CTCCGCGCGGAGACGCGC-TGCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGT 644

Qy        447 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 506
              || ||    |  | |     || | ||||  ||||||    |  ||  ||| | |  | |
Db        643 GCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGGCCTCGT 584

Qy        507 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT---GCTTGCTGAGGGGACAGTAGG 563
              || ||    |  | | |   || | ||||   ||||   |||    | || |||     |
Db        583 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTG 524

Qy        564 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 623
              || || |   ||  || |  ||   ||  | |     || ||       ||| |   |
Db        523 CCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGCATGGCCTCCAGGCGGGCCG 464

Qy        624 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 683
              || || |     || | |  | ||  | ||    ||| | || || |  || |||   |
Db        463 CCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-GCCTCCTCGAGGGCCTGGACG 405

Qy        684 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 743
              || |   |||  | |  ||  |    ||| |   | || || |     |||| |
Db        404 CGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCGCCTCCTCGGCTTCCTCCAGC 345

Qy        744 GCTGC---TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 800
              | ||    ||  | |||||     |   |  ||  || || ||||| |     |||
Db        344 TCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGC 285

Qy        801 TGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 857
              |||  | ||  | || |||| || |   || | |      ||  |  |   ||||
Db        284 ATGGCCTCCAGGCGCGCCGCCTCCTCGGCAGCCTCCAGCTCCGCGCGGAGACGCGCT 228
```

RESULT 76
AAZ25621/c
ID AAZ25621 standard; DNA; 1771 BP.
XX
AC AAZ25621;
XX
DT 22-DEC-1999 (first entry)
XX
DE Leishmania tropica Lt210 nucleotide sequence.
XX
KW Leishmania tropica; Lt210; infection; diagnosis; detection; vaccine;

```
KW   leishmaniasis; immune response; ss.
XX
OS   Leishmania tropica.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1701
FT                   /*tag= a
FT                   /product= "Lt-210"
XX
PN   US5965142-A.
XX
PD   12-OCT-1999.
XX
PF   04-AUG-1995;   95US-0511872.
XX
PR   04-AUG-1995;   95US-0511872.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Dillon DC,  Reed SG;
XX
DR   WPI; 1999-600513/51.
DR   P-PSDB; AAY45167.
XX
PT   Lt210 polypeptide useful for diagnosis and prevention of Leishmania
PT   infections -
XX
PS   Example 1; Fig 1A; 29pp; English.
XX
CC   The present sequence encodes Leishmania tropica Lt210 protein. The Lt210
CC   amino acid repeat sequence has the following 33 amino acid sequence:
CC   Gln-Ala-Arg-Xaa-Gln-Ala-Leu-Glu-Glu-Ala-Ala-Arg-Leu-Xbb-Ala-Glu-Leu-Glu-
CC   Xcc-Xcc-Glu-Glu-Ala-Ala-Arg-Leu-Xdd-Xee-Met-His-Xcc-Xff-Glu (II)
CC   Xaa = Val or Ser; Xbb = Arg or Cys; Xcc = Ala or Glu; Xdd = Asp or Glu;
CC   Xee = Val or Ala; and Xff = Ala or Gly. Proteins comprising at least one
CC   copy of the repeat sequence are useful for the diagnosis of Leishmania
CC   infections in patients and in blood supplies. The diagnosis is carried
CC   out by immunogenic assay e.g. ELISA using sera or antibodies from a
CC   patient or experimental animal or delayed-type hypersensitivity (DTH)
CC   reaction. This detection can be used to evaluate a patients capability to
CC   generate an immune response to Leishmania. These proteins can also be
CC   used as vaccines and in compositions to protect against Leishmania
CC   infections.
XX
SQ   Sequence 1771 BP; 279 A; 552 C; 672 G; 268 T; 0 other;

  Query Match             6.4%;  Score 64.2;  DB 20;  Length 1771;
  Best Local Similarity   48.7%;  Pred. No. 0.0018;
  Matches  320;  Conservative    0;  Mismatches  328;  Indels    9;  Gaps    5;

Qy        207 GACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 266
Db        881 GCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGTGCATGACATCCAGGCGGGCC 822

Qy        267 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 326
Db        821 GCCTCCTCGGCCGCCT-CCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGAC 763

Qy        327 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 386
Db        762 GCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAG 703

Qy        387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 446
Db        702 CTCCGCGCGGAGACGCGC-TGCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGT 644

Qy        447 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 506
Db        643 GCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGGCCTCGT 584

Qy        507 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT---GCTTGCTGAGGGGACAGTAGG 563
Db        583 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTG 524

Qy        564 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 623
Db        523 CCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGCATGGCCTCCAGGCGGGCCG 464

Qy        624 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 683
Db        463 CCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-GCCTCCTCGAGGGCCTGGACG 405

Qy        684 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 743
Db        404 CGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCGCCTCCTCGGCTTCCTCCAGC 345

Qy        744 GCTGC---TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 800
Db        344 TCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGC 285

Qy        801 TGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 857
Db        284 ATGGCCTCCAGGCGCGCCGCCTCCTCGGCAGCCTCCAGCTCCGCGCGGAGACGCGCT 228


RESULT 77
AAF88524/c
ID   AAF88524 standard; DNA; 1771 BP.
XX
AC   AAF88524;
XX
DT   18-NOV-2002  (first entry)
XX
DE   L. tropica LT-210 antigen DNA SEQ ID 7.
XX
KW   Antigen; immunogenic; antiparasitic; immunostimulant; leishmaniasis;
KW   gene therapy; vaccine; interleukin-12 agonist; ds.
XX
OS   Leishmania tropica.
XX
PN   US2002081320-A1.
XX
PD   27-JUN-2002.
XX
PF   04-JUN-2001; 2001US-0874923.
XX
PR   22-SEP-1995;   95US-0533669.
PR   12-FEB-1997;   97US-0798841.
PR   27-AUG-1997;   97US-0920609.
PR   12-FEB-1998;   98US-0022765.
PR   30-OCT-1998;   98US-0183861.
PR   14-APR-2000; 2000US-0551974.
PR   05-MAY-2000; 2000US-0565501.
PR   14-AUG-2000; 2000US-0639206.
XX
PA   (REED/) REED S G.
PA   (CAMP/) CAMPOS-NETO A.
PA   (WEBB/) WEBB J R.
PA   (DILL/) DILLON D C.
PA   (SKEI/) SKEIKY Y A W.
PA   (BHAT/) BHATIA A.
PA   (COLE/) COLER R N.
PA   (PROB/) PROBST P.
PA   (BRAN/) BRANNON M.
XX
PI   Reed SG,  Campos-Neto A,  Webb JR,  Dillon DC,  Skeiky YAW,  Bhatia A;
PI   Coler RN,  Probst P,  Brannon M;
XX
DR   WPI; 2002-635457/68.
DR   P-PSDB; AAB71262.
XX
PT   New polypeptide containing at least an immunogenic portion of one or
PT   more Leishmania antigens or their variants, useful for preventing,
PT   treating and detecting leishmaniasis, and stimulating immune responses
PT   in patients  -
XX
```

```
PS   Example 4; Page 34-35; 163pp; English.
XX
CC   This invention describes a novel polypeptide containing an immunogenic
CC   portion of a Leishmania antigen or its variant which has antiparasitic
CC   and immunostimulant activity. The compositions and methods of the present
CC   invention are useful for preventing, treating and detecting
CC   leishmaniasis, and stimulating immune responses in patients against
CC   leishmaniasis. The polypeptides and the polynucleotides encoding them can
CC   be used for gene therapy, in vaccines or as interleukin-12 agonists. The
CC   compositions and methods of the present invention, as compared to prior
CC   art, are more improved therapeutic modalities in the diagnosis,
CC   prevention and treatment of leishmaniasis.
XX
SQ   Sequence 1771 BP; 279 A; 552 C; 672 G; 268 T; 0 other;

  Query Match             6.4%;  Score 64.2;  DB 24;  Length 1771;
  Best Local Similarity   48.7%;  Pred. No. 0.0018;
  Matches  320;  Conservative    0;  Mismatches  328;  Indels    9;  Gaps    5;

Qy        207 GACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 266
              | | ||  |   | |||    |  ||  ||| ||| | |||        |||  || |
Db        881 GCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGTGCATGACATCCAGGCGGGCC 822

Qy        267 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 326
              || | ||||   |||| |  |||  | |   ||  |  |||| | ||   ||| | | |
Db        821 GCCTCCTCGGCCGCCT-CCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGAC 763

Qy        327 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 386
              ||  || ||||||    | ||| ||       | | |   | || || |   ||    ||
Db        762 GCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAG 703

Qy        387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 446
                | || ||  | |  || ||| | ||||  ||||||   ||  ||  ||| | |  | |
Db        702 CTCCGCGCGGAGACGCGC-TGCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGT 644

Qy        447 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 506
              || ||    |  | |     || | ||||  ||||||    |  ||  ||| | |  | |
Db        643 GCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGGCCTCGT 584

Qy        507 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT---GCTTGCTGAGGGGACAGTAGG 563
              || ||    |  | | |   || | ||||   ||||   |||    | || |||     |
Db        583 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTG 524

Qy        564 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 623
              || || |   ||   || |  ||   ||  | |     || ||       ||| |   |
Db        523 CCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGCATGGCCTCCAGGCGGGCCG 464

Qy        624 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 683
              || || |     || | |  | ||   | ||     ||| | || || |  || |||   |
Db        463 CCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-GCCTCCTCGAGGGCCTGGACG 405

Qy        684 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 743
                ||  |   |||  |  |  ||  |    ||| |   | || || |     |||| |
Db        404 CGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCGCCTCCTCGGCTTCCTCCAGC 345

Qy        744 GCTGC---TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 800
              | ||    ||  | |||||      |   |   ||  || || ||||| |      |||
Db        344 TCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGC 285

Qy        801 TGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 857
                |||  | ||  | || |||| || |   || | |      ||  |  |   ||||
Db        284 ATGGCCTCCAGGCGCGCCGCCTCCTCGGCAGCCTCCAGCTCCGCGCGGAGACGCGCT 228


RESULT 78
AAD40285/c
ID   AAD40285 standard; DNA; 1771 BP.
XX
AC   AAD40285;
XX
DT   22-OCT-2002  (first entry)
XX
DE   Leishmania tropica Lt-1 antigenic protein encoding DNA.
XX
KW   Immunogen; Leishmania antigen; therapy; delayed-type hypersensitivity;
KW   leishmaniasis; vaccine; interleukin-12 stimulation; cancer; protozoacide;
KW   virucide; bactericide; cytostatic; immune response; Lt-1 protein; gene;
KW   ds.
XX
OS   Leishmania tropica.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1671
FT                   /*tag=  a
FT                   /product= "Lt-1 antigenic protein"
FT                   /note= "No start codon"
FT                   /partial
XX
PN   US6375955-B1.
XX
PD   23-APR-2002.
XX
PF   12-FEB-1998;   98US-0022765.
XX
PR   22-SEP-1995;   95US-0533669.
PR   12-FEB-1997;   97US-0798841.
PR   27-AUG-1997;   97US-0920609.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Reed SG,  Campos-Neto A,  Webb JR,  Dillon DC,  Skeiky YAW,  Bhatia A;
PI   Probst P;
XX
DR   WPI; 2002-424754/45.
DR   P-PSDB; AAE24929.
XX
PT   New polypeptide comprising immunogenic part of Leishmania antigen,
PT   useful for treatment, prevention and diagnosis of leishmaniasis -
XX
PS   Example 4; Column 59-63; 111pp; English.
XX
CC   The present invention relates to novel proteins comprising immunogenic
CC   part of Leishmania antigen. Compositions containing sequences of the
CC   invention and other Leishmania antigen related polypeptides are useful
CC   for preventing, treating and detecting (in delayed-type hypersensitivity
CC   skin tests) leishmaniasis. They can also be used to treat any diseases
CC   responsive to interleukin-12 stimulation, including bacterial, viral and
CC   protozoal infections and cancer. Sequences of the invention are useful
CC   as vaccines. The present sequence is a DNA encoding L. tropica Lt-1
CC   (Lt-210) antigenic protein.
XX
SQ   Sequence 1771 BP; 279 A; 552 C; 672 G; 268 T; 0 other;

  Query Match             6.4%;  Score 64.2;  DB 24;  Length 1771;
  Best Local Similarity   48.7%;  Pred. No. 0.0018;
  Matches  320;  Conservative    0;  Mismatches  328;  Indels    9;  Gaps    5;

Qy        207 GACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 266
              | | ||  |   | |||    |  ||  ||| ||| | |||        |||  || |
Db        881 GCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGTGCATGACATCCAGGCGGGCC 822

Qy        267 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 326
              || | ||||   |||| |  |||  | |   ||  |  |||| | ||   ||| | | |
Db        821 GCCTCCTCGGCCGCCT-CCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGAC 763

Qy        327 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 386
              ||  || ||||||    | ||| ||       | | |   | || || |   ||    ||
Db        762 GCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAG 703

Qy        387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 446
                | || ||  | |  || ||| | ||||  ||||||   ||  ||  ||| | |  | |
Db        702 CTCCGCGCGGAGACGCGC-TGCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGT 644

Qy        447 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 506
              || ||    |  | |     || | ||||  ||||||    |  ||  ||| | |  | |
Db        643 GCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGGCCTCGT 584
```

```
Qy        507 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT---GCTTGCTGAGGGGACAGTAGG 563
              || ||    |  | | |   || | ||||  ||||   |||    | || |||     |
Db        583 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTG 524

Qy        564 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 623
               || || |   ||   || |  ||   ||  | |     || ||       ||| |   |
Db        523 CCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGCATGGCCTCCAGGCGGGCCG 464

Qy        624 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 683
               || || |     || | |  | ||   ||| | || || |     |||  |   |
Db        463 CCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-GCCTCCTCGAGGGCCTGGACG 405

Qy        684 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 743
                 ||  |   |||  |  |  ||  |   ||| |   | || || |     |||| |
Db        404 CGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCGCCTCCTCGGCTTCCTCCAGC 345

Qy        744 GCTGC---TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 800
               | ||    ||  | |||||      |   |   ||  || || ||||| |      |||
Db        344 TCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGC 285

Qy        801 TGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 857
               |||  | ||  | || |||| || |   || | |      ||  |  |   ||||
Db        284 ATGGCCTCCAGGCGCGCCGCCTCCTCGGCAGCCTCCAGCTCCGCGCGGAGACGCGCT 228
```

RESULT 79
ABK81733/c
ID ABK81733 standard; DNA; 1771 BP.
XX
AC ABK81733;
XX
DT 13-AUG-2002 (first entry)
XX
DE Leishmania antigenic polynucleotide #4.
XX
KW Leishmaniasis; protozoacide; vaccine; immune response enhancer; gene; ds;
KW Th1; interleukin-12; IL-12.
XX
OS Leishmania tropica.
XX
PN US6365165-B1.
XX
PD 02-APR-2002.
XX
PF 30-OCT-1998; 98US-0183861.
XX
PR 22-SEP-1995; 95US-0533669.
PR 12-FEB-1997; 97US-0798841.
PR 27-AUG-1997; 97US-0920609.
PR 12-FEB-1998; 98US-0022765.
XX
PA (CORI-) CORIXA CORP.
XX
PI Reed SG, Campos-Neto A, Webb JR, Dillon DC, Skeiky YAW;
XX
DR WPI; 2002-433419/46.
DR P-PSDB; ABG60867.
XX
PT Stimulating immune responses in a patient for preventing or treating
PT Leishmaniasis, by administering a vaccine comprising a polypeptide
PT comprising an immunogenic portion or epitope of Leishmania antigen -
XX
PS Example 4; Column 61-66; 114pp; English.
XX
CC The invention relates to a method for stimulating an immune response in a
CC patient comprising administering to the patient a vaccine containing a
CC non-specific immune response enhancer and a polypeptide having an
CC immunogenic portion of a Leishmania antigen, or a polypeptide having two
CC contiguous epitopes of a Leishmania antigen. The method is useful for
CC stimulating an immune response, in particular a Th1 response or
CC interleukin-12 (IL-12) production in a patient, for preventing and
CC treating Leishmaniasis. This sequence represents DNA encoding a
CC Leishmania antigenic polypeptide of the invention.
XX
SQ Sequence 1771 BP; 279 A; 552 C; 672 G; 268 T; 0 other;

```
  Query Match            6.4%;  Score 64.2;  DB 24;  Length 1771;
  Best Local Similarity  48.7%;  Pred. No. 0.0018;
  Matches  320;  Conservative    0;  Mismatches  328;  Indels    9;  Gaps    5;

Qy        207 GACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 266
              | | ||  |   | |||    |  ||  ||| |||  | |||        |||  || |
Db        881 GCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGTGCATGACATCCAGGCGGGCC 822

Qy        267 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 326
              || | ||||  |||| |  |||  | |   ||  |  |||| | ||   ||| | | |
Db        821 GCCTCCTCGGCCGCCT-CCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGAC 763

Qy        327 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 386
              ||  || ||||||    | ||| ||       | | |   | || || |   ||    ||
Db        762 GCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAG 703

Qy        387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 446
                | || ||  | |  || ||| | ||||  ||||||   ||  ||. |||.| |  | |
Db        702 CTCCGCGCGGAGACGCGC-TGCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGT 644

Qy        447 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 506
              || ||    |  | |    || | ||||  ||||||    |  ||  ||| | |  | |
Db        643 GCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGGCCTCGT 584

Qy        507 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT---GCTTGCTGAGGGGACAGTAGG 563
              || ||    |  | | |   || | ||||  ||||   |||    | || |||     |
Db        583 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTG 524

Qy        564 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 623
               || || |   ||   || |  ||   ||  | |     || ||       ||| |   |
Db        523 CCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGCATGGCCTCCAGGCGGGCCG 464

Qy        624 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 683
               || || |     || | |  | ||   ||| | || || |     |||  |   |
Db        463 CCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-GCCTCCTCGAGGGCCTGGACG 405

Qy        684 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 743
                 ||  |   |||  |  |  ||  |   ||| |   | || || |     |||| |
Db        404 CGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCGCCTCCTCGGCTTCCTCCAGC 345

Qy        744 GCTGC---TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 800
               | ||    ||  | |||||      |   |   ||  || || ||||| |      |||
Db        344 TCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGC 285

Qy        801 TGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 857
               |||  | ||  | || |||| || |   || | |      ||  |  |   ||||
Db        284 ATGGCCTCCAGGCGCGCCGCCTCCTCGGCAGCCTCCAGCTCCGCGCGGAGACGCGCT 228
```

RESULT 80
AAS96022/c
ID AAS96022 standard; cDNA; 1771 BP.
XX
AC AAS96022;
XX
DT 26-FEB-2002 (first entry)
XX
DE Leishmania antigen Lt-1 DNA.
XX
KW Leishmaniasis; Leishmania antigen; immunostimulant; protozoacide; Ldp23;
KW interleukin-15; PCR primer; Lbhsp83; M15; Lt-1; LbeIF4A; Lmsp1a; Lmsp9a;
KW MAPS-1A; LmgSP1; LmgSP3; LmgSP5; LmgSP8; LmgSP9; LmgSP13; LmgSP19; ss;
KW LcgSP1; LcgSP3; LcgSP4; LcgSP8; LcgSP10; 1G6-34; 1E6-44; 4A5-63; 1B11-39;
KW 2A10-37; 4G2-83; 4H6-41; 8G3-100.
XX
OS Leishmania tropica.
XX
PN WO200179276-A2.
XX

```
PD   25-OCT-2001.
XX
PF   05-APR-2001; 2001WO-US11254.
XX
PR   14-APR-2000; 2000US-0551974.
PR   05-MAY-2000; 2000US-0565501.
PR   14-AUG-2000; 2000US-0639206.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Reed SG,  Campos-Neto A,  Webb JR,  Dillon DC,  Skeiky YAW,  Bhatia A;
PI   Coler RM,  Probst P;
XX
DR   WPI; 2002-061971/08.
DR   P-PSDB; AAU71808.
XX
PT   New isolated Leishmania antigens, useful for prevention, treatment and
PT   diagnosis of leishmaniasis, also related nucleic acids for genetic
PT   vaccination  -
XX
PS   Example 4; Page 120-121; 193pp; English.
XX
CC   The invention relates to polypeptides comprising an immunogenic part of a
CC   Leishmania antigen. The Leishmania polypeptides and their associated DNA
CC   sequences, epitopes and fusion proteins are used in the production of
CC   compositions used for inducing a protective immune response against
CC   leishmaniasis, for prevention and treatment of the disease. The
CC   compositions can also be used generally to treat diseases that respond to
CC   interleukin-15 stimulation. In addition, the products may contain an
CC   immunostimulant. The sequences represent cDNA molecules encoding
CC   Leishmania antigens and PCR primers used to amplify DNA of the invention.
XX
SQ   Sequence 1771 BP; 279 A; 552 C; 672 G; 268 T; 0 other;

  Query Match             6.4%;  Score 64.2;  DB 24;  Length 1771;
  Best Local Similarity   48.7%;  Pred. No. 0.0018;
  Matches  320;  Conservative    0;  Mismatches  328;  Indels    9;  Gaps    5;

Qy        207 GACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 266
              | | ||  |   | |||    |  ||  ||| |||  | |||       |||  || |
Db        881 GCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGTGCATGACATCCAGGCGGGCC 822

Qy        267 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 326
              || | ||||   |||| |  |||  | |   ||  |  |||| | ||   ||| | | |
Db        821 GCCTCCTCGGCCGCCT-CCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGAC 763

Qy        327 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 386
              ||  || ||||||    | ||| ||       | | |   | || || |   ||    ||
Db        762 GCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAG 703

Qy        387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 446
                | || ||  | |  || ||| | ||||  ||||||   ||  ||  ||| | |  | |
Db        702 CTCCGCGCGGAGACGCGC-TGCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGT 644

Qy        447 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 506
              || ||    |  | |     || | ||||  ||||||    |  ||  ||| | |  | |
Db        643 GCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGGCCTCGT 584

Qy        507 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT---GCTTGCTGAGGGGACAGTAGG 563
              || ||    |  | | |   || | ||||   ||||   |||    | || |||     |
Db        583 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTG 524

Qy        564 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 623
              || || |   ||   || |  ||   ||  | |     || ||        ||| |   |
Db        523 CCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGCATGGCCTCCAGGCGGGCCG 464

Qy        624 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 683
              || || |     || | |  | ||  | ||    ||| | || || |  || |||   |
Db        463 CCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-GCCTCCTCGAGGGCCTGGACG 405

Qy        684 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 743
                || |  |||  | |  ||  |    ||| |   | || || |     |||| |
Db        404 CGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCGCCTCCTCGGCTTCCTCCAGC 345

Qy        744 GCTGC---TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 800
              | ||    ||  | |||||      |   |   ||  || || ||||| |      |||
Db        344 TCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGC 285

Qy        801 TGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 857
                |||  | ||  | || |||| || |   || | |      ||  |  |   ||||
Db        284 ATGGCCTCCAGGCGCGCCGCCTCCTCGGCAGCCTCCAGCTCCGCGCGGAGACGCGCT 228


RESULT 81
AAL44951
ID   AAL44951 standard; DNA; 8174 BP.
XX
AC   AAL44951;
XX
DT   13-MAY-2002  (first entry)
XX
DE   Equine herpesvirus type 1 immediate early gene.
XX
KW   Equine herpes virus; immediate early; vaccine; EHV-1; virucide;
KW   immunostimulant; gene; ds.
XX
OS   Equine herpesvirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             988..5451
FT                   /*tag= a
FT                   /product= "immediate early protein"
XX
PN   WO200209750-A2.
XX
PD   07-FEB-2002.
XX
PF   27-JUL-2001; 2001WO-US23612.
XX
PR   27-JUL-2000; 2000US-0626748.
XX
PA   (RESE ) RESEARCH CORP TECHNOLOGIES INC.
XX
PI   O'Callaghan DJ;
XX
DR   WPI; 2002-206153/26.
DR   P-PSDB; AAO14351.
XX
PT   Novel mutant equine herpesvirus type-1 isolates having mutation in
PT   immediate early gene, useful in formulating vaccine compositions for
PT   preventing and treating equine herpesvirus type-1 infections in horses
PT   -
XX
PS   Disclosure; Page 47-51; 68pp; English.
XX
CC   The present invention relates to a mutant equine herpesvirus type-1
CC   (EHV-1) isolate, in particular a replication-competent EHV-1 isolate
CC   comprising a mutation in the immediate-early (IE) gene of the viral
CC   genome. The mutant isolate is useful for stimulating an immune response,
CC   preferably a cell-mediated or humoral immune response, against EHV-1,
CC   and for preventing or inhibiting an EHV-1 infection in a horse. The
CC   present sequence is the EHV-1 immediate early gene.
XX
SQ   Sequence 8174 BP; 1171 A; 2937 C; 2790 G; 1276 T; 0 other;

  Query Match             6.3%;  Score 63.4;  DB 24;  Length 8174;
  Best Local Similarity   50.8%;  Pred. No. 0.0027;
  Matches  151;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy        445 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 504
              ||| || || |  ||  || | |   | | | ||| | | ||| | |   | | |   |
Db       5774 TTGTTTACTGGGCGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGG 5833

Qy        505 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGG 564
                | ||| | |  ||  || | |   | | | ||| | | ||| || | ||   | |||
Db       5834 CGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGG 5893
```

```
Qy        565 CTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGG 624
Db       5894 CGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGG 5953

Qy        625 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 684
Db       5954 CGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGGCGGGTTGATGGG 6013

Qy        685 CTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCT 741
Db       6014 CGGGTTGATGGTTCCTGCTCCTCCCCTTCCTGCTCCTCCCCTTCCTGCTCCTCCCCT 6070
```

RESULT 82
AAZ17263/c
ID AAZ17263 standard; cDNA; 1337 BP.
XX
AC AAZ17263;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:4735.
XX
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9938972-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01619.
XX
PR 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI; 1999-494092/41.
XX
PT Novel human genes and their expression products which are
PT differentially expressed in different cell types
XX
PS Claim 1; Page 2250-2251; 2479pp; English.
XX
CC The present invention describes a library of human polynucleotides
CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
SQ Sequence 1337 BP; 42 A; 577 C; 27 G; 22 T; 669 other;

Query Match 6.3%; Score 63; DB 20; Length 1337;
Best Local Similarity 24.4%; Pred. No. 0.0031;
Matches 221; Conservative 0; Mismatches 682; Indels 1; Gaps 1;

```
Qy         57 GGTGGTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGC 116
Db       1256 GNGNGNNNGNNNGNNNGNGGGNGGGNGNGGGNGNNNNNNNGNGNGGNGNGNGNNGNGGNN 1197

Qy        117 GCGGGGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGG 176
Db       1196 NGNNNGNGNNNGNGNGNGNGNGGNGNNNNGNNNNNGNNGNNNNNNGNGGNGNGNNNN 1137

Qy        177 GGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGC 236
Db       1136 NNGGNGNNGNNGNNGGNNGGGGNGNGNGNGNGNGNNNNGNNNGNNG-NNGGNNGNNNGN 1078

Qy        237 TGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 296
Db       1077 GNGNNGNGGGNGAGNNNNNNGNGGNGGNGGNNGGNNGNGGGGNNNNNNNNNGNNGNNGNN 1018

Qy        297 CAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAG 356
Db       1017 NNNGGNNNNNGNGNGNNGNNGGNGNNGNGNNNGGNNNNGNNNNGNGNNNNGNGNNNNG 958

Qy        357 GGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGC 416
Db        957 GNGNNNNNNGGNNGNGNNGNNNNNGNGNNGGNNNGNGNNNGGNGNNGGNGGGGNNNGNNG 898

Qy        417 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 476
Db        897 NNGNNNGGGGNNNGGNNGGNGGNNGNGGNNNGGNGNGNNNGGGNGNGNNNNNNGGNGGGG 838

Qy        477 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 536
Db        837 NGGNGNGNNNNGGNGGNGNGNGGGGNGNNGNNNNGNGGNGGGGNNNNNNNGNGNGGNGGN 778

Qy        537 TGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGC 596
Db        777 NGGNNGGNGGNNGGGNGNNGGNGGGNNGGGGGNGGGGNNNGGGNNGGGGGGNGGGGGGNN 718

Qy        597 CTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 656
Db        717 GNGGNGGNGGGGGGNGGGGGGGGGNGGGNGNNGGNNGGGGGGGGNGGNGNNGGGGGGGGG 658

Qy        657 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGG 716
Db        657 GGGGGGGGGGGNNNNGGGGGGGGGGGGGNGNGGNGGGGGGNNNGNGGGGGGGGNGGGGGG 598

Qy        717 CTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 776
Db        597 NGNGNGGNNGGGGGGGNGGNGGGGGGNNNNGGNGGGGNNGGGNNNGGGGGNGNGNGNGGG 538

Qy        777 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGC 836
Db        537 GNGGGNNGGGGNNGGGNGGGGGGGGGGGGGGGGNGNGGGGNGGGGGNGNGGNNGGGGNNG 478

Qy        837 TTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGA 896
Db        477 NNNNNNNNGGGGGGGGNNNNGNNNNNGNGGGGGNGGNNNNNNGGGNGNNNGGGGNGNNGG 418

Qy        897 GTAGGGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACC 956
Db        417 NGNGGGGGGNNNNNNNNNGNGNNNNNGNGGNNNNGGGNGNNNNGNNGNNNNNGNNNNNN 358
```

```
Qy        957 CGGG 960
              | |
Db        357 NGNG 354


RESULT 83
AAL40260/c
ID   AAL40260 standard; DNA; 149 BP.
XX
AC   AAL40260;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 9.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   P-PSDB; AAO21907.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear  -
XX
PS   Example 2; Fig 3; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 149 BP; 62 A; 31 C; 26 G; 30 T; 0 other;

  Query Match             6.3%;  Score 62.8;  DB 24;  Length 149;
  Best Local Similarity   85.4%;  Pred. No. 0.0032;
  Matches   70;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        389 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 448
              ||||| ||| || |||||||||||| ||| || |||||||||||| ||| || |||||||
Db         92 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 33

Qy        449 TTGCTCGCTGGCCTGCTTGCTT 470
              ||||| ||| || |||||||||
Db         32 TTGCTTGCTTGCTTGCTTGCTT 11


RESULT 84
AAL40261/c
ID   AAL40261 standard; DNA; 149 BP.
XX
AC   AAL40261;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 11.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   P-PSDB; AAO21908.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear  -
XX
PS   Example 2; Fig 3; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
```

```
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 149 BP; 62 A; 31 C; 26 G; 30 T; 0 other;

  Query Match             6.3%;  Score 62.8;  DB 24;  Length 149;
  Best Local Similarity   85.4%;  Pred. No. 0.0032;
  Matches   70;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        389 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 448
              ||||| ||| || ||||||||||||| ||| || ||||||||||||| ||| || |||||||
Db         92 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 33

Qy        449 TTGCTCGCTGGCCTGCTTGCTT 470
              ||||| ||| || |||||||||
Db         32 TTGCTTGCTTGCTTGCTTGCTT 11


RESULT 85
AAL40262/c
ID   AAL40262 standard; DNA; 149 BP.
XX
AC   AAL40262;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 13.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   P-PSDB; AAO21909.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear -
XX
PS   Example 2; Fig 3; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 149 BP; 62 A; 31 C; 26 G; 30 T; 0 other;

  Query Match             6.3%;  Score 62.8;  DB 24;  Length 149;
  Best Local Similarity   85.4%;  Pred. No. 0.0032;
  Matches   70;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        389 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 448
              ||||| ||| || ||||||||||||| ||| || ||||||||||||| ||| || |||||||
Db         92 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 33

Qy        449 TTGCTCGCTGGCCTGCTTGCTT 470
              ||||| ||| || |||||||||
Db         32 TTGCTTGCTTGCTTGCTTGCTT 11


RESULT 86
AAL40263/c
ID   AAL40263 standard; DNA; 149 BP.
XX
AC   AAL40263;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 15.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   P-PSDB; AAO21910.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear -
XX
PS   Example 2; Fig 3; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
```

```
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 149 BP; 62 A; 31 C; 26 G; 30 T; 0 other;

  Query Match             6.3%;  Score 62.8;  DB 24;  Length 149;
  Best Local Similarity   85.4%;  Pred. No. 0.0032;
  Matches   70;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        389 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 448
              ||||| ||| || |||||||||||| ||| || |||||||||||| ||| || |||||||
Db         92 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 33

Qy        449 TTGCTCGCTGGCCTGCTTGCTT 470
              ||||| ||| || |||||||||
Db         32 TTGCTTGCTTGCTTGCTTGCTT 11


RESULT 87
AAL40256/c
ID   AAL40256 standard; DNA; 2084 BP.
XX
AC   AAL40256;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 1.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
DR   N-PSDB; AAO21901, AAO21902.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear  -
XX
PS   Example 1; Fig 1; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 2084 BP; 706 A; 444 C; 400 G; 534 T; 0 other;

  Query Match             6.3%;  Score 62.8;  DB 24;  Length 2084;
  Best Local Similarity   85.4%;  Pred. No. 0.0034;
  Matches   70;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        389 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 448
              ||||| ||| || |||||||||||| ||| || |||||||||||| ||| || |||||||
Db         92 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 33

Qy        449 TTGCTCGCTGGCCTGCTTGCTT 470
              ||||| ||| || |||||||||
Db         32 TTGCTTGCTTGCTTGCTTGCTT 11


RESULT 88
AAL40257/c
ID   AAL40257 standard; DNA; 2084 BP.
XX
AC   AAL40257;
XX
DT   13-SEP-2002  (first entry)
XX
DE   BASB204 related DNA sequence SEQ ID No 3.
XX
KW   Antibacterial; auditory; antiinflammatory; antiviral; BASB204 protein;
KW   Haemophilus influenzae; immunogenic; immunospecific; Gram-negative;
KW   Gram-positive bacteria; mammalian extracellular matrix protein; wound;
KW   bacterial adhesion; tissue damage; pathogenesis; infection; otitis media;
KW   pneumonia; sinusitis; nosocomial infection; invasive disease; middle ear;
KW   hearing loss; chronic otitis media; fluid accumulation; inflammation;
KW   auditive nerve damage; delayed speech learning; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200234768-A2.
XX
PD   02-MAY-2002.
XX
PF   24-OCT-2001; 2001WO-EP12390.
XX
PR   24-OCT-2000; 2000GB-0026002.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2002-471431/50.
```

```
DR   N-PSDB; AAO21903, AAO21904.
XX
PT   New isolated non-typeable Haemophilus influenzae BASB204 polypeptides
PT   useful as components of vaccines useful for treating bacterial
PT   infection such as otitis media, delayed speech learning, and
PT   inflammation of the middle ear -
XX
PS   Example 1; Fig 1; 97pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino acid
CC   sequence which has 85% identity to the entire length of a fully defined
CC   non-typeable Haemophilus influenzae BASB204 polypeptide sequence, given
CC   in specification. The isolated protein, an immunogenic fragment of the
CC   protein or an antibody immunospecific for the protein or immunogenic
CC   fragment is useful for diagnosing a non-typeable H. influenzae infection.
CC   The molecules may be used in the prevention of adhesion of Gram-negative
CC   or Gram-positive bacteria to mammalian extracellular matrix proteins in
CC   wounds, to block bacterial adhesion between mammalian extracellular
CC   matrix and bacterial BASB204 protein that mediate tissue damage and/or to
CC   block the normal progression of pathogenesis in infections initiated
CC   other than by the implantation of in-dwelling devices. The isolated
CC   protein encoding polynucleotide is useful as a research reagent and
CC   material for discovery of treatments of and diagnostics for diseases,
CC   particularly human diseases. The antibody of the invention is useful for
CC   treating infections preferably bacterial infections such as otitis media
CC   in infants and children, pneumonia in elders, sinusitis, nosocomial
CC   infections, or invasive diseases, chronic otitis media with hearing loss,
CC   fluid accumulation in the middle ear, auditive nerve damage, delayed
CC   speech learning, infections of the upper respiratory tract and
CC   inflammation of the middle ear. This polynucleotide sequence represents a
CC   Haemophilus influenzae BASB204 related nucleic acid of the invention.
XX
SQ   Sequence 2084 BP; 704 A; 444 C; 401 G; 535 T; 0 other;

  Query Match             6.3%;  Score 62.8;  DB 24;  Length 2084;
  Best Local Similarity   85.4%;  Pred. No. 0.0034;
  Matches   70;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        389 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 448
              ||||| ||| || |||||||||||| ||| || |||||||||||| ||| || |||||||
Db         92 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 33

Qy        449 TTGCTCGCTGGCCTGCTTGCTT 470
              ||||| ||| || |||||||||
Db         32 TTGCTTGCTTGCTTGCTTGCTT 11


RESULT 89
AAS39915
ID   AAS39915 standard; DNA; 31718 BP.
XX
AC   AAS39915;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Genomic sequence #334 encoding human colon associated polypeptide.
XX
KW   Human; colon cancer; congenital abnormality; infection; colitis;
KW   inflammatory bowel disease; IBD; neoplastic disorder; gene therapy;
KW   intestinal inflammatory disorder; malabsorption syndrome; gastric;
KW   sigmoid disease; antibacterial; antiviral; antiinflammatory;
KW   cytostatic; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155302-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01240.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
```

PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465567/50.
XX
PT Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders related to the colon including colon cancers and also for
PT testing and detection e.g. diagnosis -
XX
PS Disclosure; SEQ ID No 812; 562pp; English.
XX
CC The present invention relates to the isolation of novel human colon
CC associated polypeptides (AAU22468-AAU22701), and the cDNA and genomic
CC sequences encoding for them. The sequences of the invention are useful
CC in the diagnosis, treatment, prevention and/or prognosis of disorders
CC of the colon including colon cancer, congenital abnormalities
CC (e.g. atresia and stenosis), bacterial and viral infections,
CC inflammatory bowel disease (IBD), neoplastic cell disorders
CC (e.g. polyps and adenomas, intestinal inflammatory disorders, colitis,
CC colonic inflammation, diarrhoea and dysentery, malabsorption syndromes
CC (e.g. lactose intolerance), intestinal obstruction and sigmoid diseases.
CC The polynucleotides sequences of the invention can also be used in gene
CC therapy. AAS39582-AAS40060 represent DNA sequences encoding for the
CC novel human colon associated polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 31718 BP; 7951 A; 7687 C; 8287 G; 7793 T; 0 other;

Query Match 6.3%; Score 62.8; DB 22; Length 31718;
Best Local Similarity 69.7%; Pred. No. 0.0037;
Matches 85; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455
Db 14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy 456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515
Db 14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy 516 CT 517
Db 15036 CT 15037

RESULT 90
AAS39916
ID AAS39916 standard; DNA; 31718 BP.
XX
AC AAS39916;
XX
DT 17-DEC-2001 (first entry)
XX
DE Genomic sequence #335 encoding human colon associated polypeptide.
XX
KW Human; colon cancer; congenital abnormality; infection; colitis;
KW inflammatory bowel disease; IBD; neoplastic disorder; gene therapy;
KW intestinal inflammatory disorder; malabsorption syndrome; gastric;
KW sigmoid disease; antibacterial; antiviral; antiinflammatory;
KW cytostatic; ds.
XX
OS Homo sapiens.
XX
PN WO200155302-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01240.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465567/50.
XX
PT Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders related to the colon including colon cancers and also for
PT testing and detection e.g. diagnosis -
XX
PS Disclosure; SEQ ID No 813; 562pp; English.
XX
CC The present invention relates to the isolation of novel human colon
CC associated polypeptides (AAU22468-AAU22701), and the cDNA and genomic
CC sequences encoding for them. The sequences of the invention are useful
CC in the diagnosis, treatment, prevention and/or prognosis of disorders
CC of the colon including colon cancer, congenital abnormalities
CC (e.g. atresia and stenosis), bacterial and viral infections,
CC inflammatory bowel disease (IBD), neoplastic cell disorders
CC (e.g. polyps and adenomas, intestinal inflammatory disorders, colitis,
CC colonic inflammation, diarrhoea and dysentery, malabsorption syndromes

```
CC   (e.g. lactose intolerance), intestinal obstruction and sigmoid diseases.
CC   The polynucleotides sequences of the invention can also be used in gene
CC   therapy. AAS39582-AAS40060 represent DNA sequences encoding for the
CC   novel human colon associated polypeptides of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 31718 BP; 7949 A; 7690 C; 8288 G; 7791 T; 0 other;

  Query Match             6.3%;  Score 62.8;  DB 22;  Length 31718;
  Best Local Similarity   69.7%;  Pred. No. 0.0037;
  Matches   85;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy        396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455
              ||  | | ||||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy        456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy        516 CT 517
              ||
Db      15036 CT 15037


RESULT 91
AAL06415
ID   AAL06415 standard; DNA; 31718 BP.
XX
AC   AAL06415;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 9103.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
```

PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; SEQ ID NO 9103; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 31718 BP; 7951 A; 7687 C; 8287 G; 7793 T; 0 other;

Query Match 6.3%; Score 62.8; DB 22; Length 31718;
Best Local Similarity 69.7%; Pred. No. 0.0037;
Matches 85; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```
Qy      396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455
            ||  | | ||||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db    14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy      456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515
            ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db    14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy      516 CT 517
            ||
Db    15036 CT 15037
```

RESULT 92
AAL06416
ID AAL06416 standard; DNA; 31718 BP.
XX
AC AAL06416;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 9104.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; SEQ ID NO 9104; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 31718 BP; 7949 A; 7690 C; 8288 G; 7791 T; 0 other;

Query Match 6.3%; Score 62.8; DB 22; Length 31718;
Best Local Similarity 69.7%; Pred. No. 0.0037;
Matches 85; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455
|| | | |||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db 14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy 456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515
|| | | ||| ||| || || | | ||| ||| ||| | | ||| ||| ||
Db 14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy 516 CT 517
||
Db 15036 CT 15037

RESULT 93
AAK73104/c
ID AAK73104 standard; DNA; 31718 BP.
XX
AC AAK73104;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27916.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 27916; 3071pp + Sequence Listing; English.

```
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 31718 BP; 7793 A; 8287 C; 7687 G; 7951 T; 0 other;

  Query Match             6.3%;  Score 62.8;  DB 22;  Length 31718;
  Best Local Similarity   69.7%;  Pred. No. 0.0037;
  Matches   85;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy        396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455
              ||  | | ||||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      16803 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 16744

Qy        456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      16743 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 16684

Qy        516 CT 517
              ||
Db      16683 CT 16682


RESULT 94
AAK73120/c
ID   AAK73120 standard; DNA; 31718 BP.
XX
AC   AAK73120;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27932.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
```

PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 27932; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 31718 BP; 7791 A; 8288 C; 7690 G; 7949 T; 0 other;

Query Match 6.3%; Score 62.8; DB 22; Length 31718;
Best Local Similarity 69.7%; Pred. No. 0.0037;
Matches 85; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455
Db 16803 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 16744

Qy 456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515
Db 16743 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 16684

Qy 516 CT 517
Db 16683 CT 16682

RESULT 95
AAK87442
ID AAK87442 standard; DNA; 31718 BP.
XX
AC AAK87442;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:42254.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 42254; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 31718 BP; 7951 A; 7687 C; 8287 G; 7793 T; 0 other;

Query Match 6.3%; Score 62.8; DB 22; Length 31718;
Best Local Similarity 69.7%; Pred. No. 0.0037;
Matches 85; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```
Qy        396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455
              ||  | | |||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy        456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy        516 CT 517
              ||
Db      15036 CT 15037


RESULT 96
AAK87443
ID   AAK87443 standard; DNA; 31718 BP.
XX
AC   AAK87443;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:42255.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
```

```
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 42255; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 31718 BP; 7949 A; 7690 C; 8288 G; 7791 T; 0 other;

  Query Match             6.3%;  Score 62.8;  DB 22;  Length 31718;
  Best Local Similarity   69.7%;  Pred. No. 0.0037;
  Matches   85;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy        396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455
              ||  | | ||||||||||| ||| || ||||||||||||| ||| || ||||||||||||
Db      14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTTCTTGCTTT 14975

Qy        456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy        516 CT 517
              ||
Db      15036 CT 15037

RESULT 97
AAK87573/c
ID   AAK87573 standard; DNA; 31718 BP.
XX
AC   AAK87573;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:42385.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
```

PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 42385; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 31718 BP; 7793 A; 8287 C; 7687 G; 7951 T; 0 other;

Query Match 6.3%; Score 62.8; DB 22; Length 31718;
Best Local Similarity 69.7%; Pred. No. 0.0037;
Matches 85; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455
|| | | |||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db 16803 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 16744

Qy 456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515
|| | | ||| ||| || || | | ||| ||| ||| | | ||| ||| ||
Db 16743 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 16684

Qy 516 CT 517
||
Db 16683 CT 16682

RESULT 98
AAK87592/c
ID AAK87592 standard; DNA; 31718 BP.
XX
AC AAK87592;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:42404.
XX

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.

PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure;. SEQ ID NO 42404; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 31718 BP; 7791 A; 8288 C; 7690 G; 7949 T; 0 other;

Query Match 6.3%; Score 62.8; DB 22; Length 31718;
Best Local Similarity 69.7%; Pred. No. 0.0037;
Matches 85; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455
Db 16803 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 16744

Qy 456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515
Db 16743 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 16684

Qy 516 CT 517
Db 16683 CT 16682

RESULT 99
AAK90359
ID AAK90359 standard; DNA; 31718 BP.
XX
AC AAK90359;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen genomic sequence SEQ ID NO: 3935.
XX
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX
OS Homo sapiens.
XX
PN WO200155314-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01324.
XX

PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases -
XX
PS Disclosure; SEQ ID NO 3935; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention.
XX
SQ Sequence 31718 BP; 7951 A; 7687 C; 8287 G; 7793 T; 0 other;

Query Match 6.3%; Score 62.8; DB 22; Length 31718;
Best Local Similarity 69.7%; Pred. No. 0.0037;
Matches 85; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```
Qy      396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455
            ||  | | ||||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db    14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy      456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515
            ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db    14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy      516 CT 517
            ||
Db    15036 CT 15037
```

RESULT 100
AAK90360
ID AAK90360 standard; DNA; 31718 BP.
XX
AC AAK90360;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen genomic sequence SEQ ID NO: 3936.
XX
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX
OS Homo sapiens.
XX
PN WO200155314-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01324.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases -
XX
PS Disclosure; SEQ ID NO 3936; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention.
XX
SQ Sequence 31718 BP; 7949 A; 7690 C; 8288 G; 7791 T; 0 other;

Query Match 6.3%; Score 62.8; DB 22; Length 31718;
Best Local Similarity 69.7%; Pred. No. 0.0037;
Matches 85; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455
|| | | |||||||||| ||| || |||||||||||| ||| || ||||||||||| .
Db 14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

```
Qy        456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy        516 CT 517
              ||
Db      15036 CT 15037


Search completed: August 21, 2003, 21:25:16
Job time : 296.141 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 21, 2003, 18:43:32 ; Search time 3722.36 Seconds
(without alignments)
11001.231 Million cell updates/sec

Title: US-09-831-000-1_COPY_132719_133719
Perfect score: 1001
Sequence: 1 cgccgcccggcatggcggtc.....cgggcacccggcccagatct 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1001	100.0	133719	14	AF083501	AF083501 Macaca mu
c	2	282.6	28.2	148106	2	AC116487	AC116487 Mus muscu
	3	262.2	26.2	265401	2	AC099177	AC099177 Rattus no
	4	257.6	25.7	165180	2	AC119239	AC119239 Mus muscu
	5	239.2	23.9	130733	14	AF210726	AF210726 Macaca mu
c	6	228	22.8	218921	2	AC097912	AC097912 Rattus no
c	7	224	22.4	229535	2	AC130110	AC130110 Rattus no
c	8	219	21.9	208050	2	AC137146	AC137146 Mus muscu
c	9	219	21.9	228919	2	AC112675	AC112675 Mus muscu
c	10	216.6	21.6	245853	2	AC131862	AC131862 Rattus no
	11	216.6	21.6	305166	2	AC106141	AC106141 Rattus no
	12	210.4	21.0	220464	2	AC115409	AC115409 Rattus no
	13	209.8	21.0	238215	2	AC141883	AC141883 Mus muscu
	14	208.6	20.8	128953	9	AC107913	AC107913 Homo sapi
	15	201	20.1	276954	2	AC098079	AC098079 Rattus no
	16	199.4	19.9	229010	2	AC073686	AC073686 Mus muscu
	17	197.8	19.8	271190	2	AC117991	AC117991 Mus muscu
	18	196.8	19.7	219713	2	AC132225	AC132225 Mus muscu
	19	195	19.5	27213	9	AP000291	AP000291 Homo sapi
	20	195	19.5	100000	9	AP000043	AP000043 Homo sapi
	21	195	19.5	100000	9	AP000111	AP000111 Homo sapi
	22	195	19.5	100000	9	AP000187	AP000187 Homo sapi
	23	195	19.5	340000	9	AP001716	AP001716 Homo sapi
c	24	191.4	19.1	133661	14	U93872	U93872 Kaposi's sa
c	25	190.2	19.0	183863	2	AC110455	AC110455 Rattus no
	26	189.8	19.0	234939	2	AC094809	AC094809 Rattus no
	27	189.2	18.9	224055	2	AC096116	AC096116 Rattus no
c	28	186.8	18.7	208740	2	AC114591	AC114591 Mus muscu
	29	185.4	18.5	209013	10	AL844846	AL844846 Mouse DNA
c	30	183.4	18.3	235855	2	AC094484	AC094484 Rattus no
	31	180.4	18.0	230500	2	AC097040	AC097040 Rattus no
c	32	180.4	18.0	258132	2	AC142075	AC142075 Rattus no
	33	179.8	18.0	110000	2	AC109525_0	AC109525 Rattus no
c	34	179.8	18.0	237019	2	AC094443	AC094443 Rattus no
c	35	178.8	17.9	189096	5	AL935202	AL935202 Zebrafish
	36	178.6	17.8	223156	2	AC112928	AC112928 Mus muscu
c	37	175.8	17.6	203152	2	AC131371	AC131371 Rattus no
c	38	175.8	17.6	270245	2	AC095180	AC095180 Rattus no
	39	175.2	17.5	201258	2	AC141889	AC141889 Mus muscu
	40	174	17.4	28199	5	AF013613	AF013613 Fugu rubr
	41	174	17.4	28199	6	AX652127	AX652127 Sequence
	42	173.4	17.3	599	9	AF388526	AF388526 Homo sapi
	43	173.2	17.3	235808	2	AC091372	AC091372 Rattus no
c	44	173.2	17.3	305226	2	AC107437	AC107437 Rattus no
	45	173	17.3	88716	2	AC098644	AC098644 Takifugu
c	46	172.4	17.2	197891	2	AC109837	AC109837 Rattus no
c	47	169.4	16.9	245258	2	AC113840	AC113840 Rattus no
c	48	169.2	16.9	201258	2	AC141889	AC141889 Mus muscu
	49	167.6	16.7	275443	2	AC096407	AC096407 Rattus no
c	50	167.4	16.7	271233	2	AC096597	AC096597 Rattus no
c	51	166.2	16.6	246968	2	AC113967	AC113967 Mus muscu
	52	164.4	16.4	234474	2	AC103472	AC103472 Rattus no
c	53	164.4	16.4	235837	2	AC112613	AC112613 Rattus no
	54	164	16.4	200141	10	AL845305	AL845305 Mouse DNA
	55	164	16.4	256110	2	AC116951	AC116951 Mus muscu
	56	162.6	16.2	197153	2	AC141966	AC141966 Rattus no
c	57	162.6	16.2	228645	2	AC094273	AC094273 Rattus no
	58	161.4	16.1	252801	2	AC095211	AC095211 Rattus no
	59	161	16.1	242482	2	AC115149	AC115149 Rattus no
c	60	161	16.1	257675	2	AC094136	AC094136 Rattus no
	61	160.6	16.0	205515	2	AC118905	AC118905 Rattus no
	62	160.6	16.0	272652	2	AC109709	AC109709 Rattus no
	63	159.2	15.9	206805	2	AC118832	AC118832 Rattus no
c	64	158	15.8	256486	2	AC106350	AC106350 Rattus no
c	65	157.6	15.7	231321	2	AC098121	AC098121 Rattus no

	66	156.8	15.7	176782	10	AC124492 AC124492 Mus muscu
c	67	156.6	15.6	140898	2	AC142424 AC142424 Rattus no
	68	156.6	15.6	179192	2	AC044848 AC044848 Mus muscu
c	69	156.6	15.6	220050	10	AL583887 AL583887 Mouse DNA
	70	156.6	15.6	236344	10	AL583889 AL583889 Mouse DNA
c	71	156.4	15.6	251656	2	AC106240 AC106240 Rattus no
c	72	155.4	15.5	208994	2	AC122561 AC122561 Mus muscu
	73	155.4	15.5	217837	10	AL772377 AL772377 Mouse DNA
	74	155.2	15.5	196519	2	AC128183 AC128183 Rattus no
	75	155	15.5	152054	10	AC122009 AC122009 Mus muscu
	76	154.2	15.4	210964	2	AC113245 AC113245 Mus muscu
	77	153.4	15.3	148066	10	AL606528 AL606528 Mouse DNA
c	78	153.2	15.3	217604	2	AC118179 AC118179 Rattus no
c	79	152.8	15.3	238112	2	AC106600 AC106600 Rattus no
c	80	152.6	15.2	201016	2	AC120242 AC120242 Rattus no
c	81	152.4	15.2	157624	10	AL669980 AL669980 Mouse DNA
c	82	152.4	15.2	262289	2	AC117951 AC117951 Rattus no
	83	151.8	15.2	223892	2	AC110123 AC110123 Rattus no
	84	151.4	15.1	243716	2	AL833804 AL833804 Mus muscu
	85	150.8	15.1	283593	2	AC131433 AC131433 Rattus no
	86	150.4	15.0	189236	2	AL645694 AL645694 Mus muscu
c	87	150.2	15.0	110000	2	AC108277_1 Continuation (2 of
c	88	150.2	15.0	110000	10	AE014176_0 AE014176 Mus muscu
c	89	150.2	15.0	189576	2	AC114585 AC114585 Mus muscu
	90	150.2	15.0	238255	2	AC132758 AC132758 Rattus no
	91	150.2	15.0	240200	2	AC116206 AC116206 Rattus no
	92	150.2	15.0	242974	2	AC103570 AC103570 Rattus no
c	93	150.2	15.0	244301	2	AC116072 AC116072 Rattus no
	94	149.8	15.0	238046	2	AC095003 AC095003 Rattus no
	95	149.2	14.9	238488	2	AC103134 AC103134 Rattus no
c	96	148.4	14.8	231764	2	AC119470 AC119470 Rattus no
	97	147.4	14.7	191953	2	AC119267 AC119267 Mus muscu
	98	146.8	14.7	165608	2	AC087842 AC087842 Rattus no
c	99	146.4	14.6	165118	2	AC119053 AC119053 Danio rer
	100	146.4	14.6	174032	2	AC128150 AC128150 Rattus no
c	101	146.4	14.6	253076	2	AC111479 AC111479 Rattus no
c	102	146.2	14.6	201290	10	AL731682 AL731682 Mouse DNA
c	103	146	14.6	160805	2	AC116801 AC116801 Mus muscu
c	104	146	14.6	172445	2	AC128287 AC128287 Rattus no
	105	145.6	14.5	176961	5	AL928917 AL928917 Zebrafish
c	106	145	14.5	240384	2	AC097041 AC097041 Rattus no
	107	144.8	14.5	223379	2	AC106435 AC106435 Rattus no
	108	144.8	14.5	265707	2	AC111671 AC111671 Rattus no
	109	144.6	14.4	187995	10	AC127254 AC127254 Mus muscu
c	110	144.6	14.4	291205	2	AC103125 AC103125 Rattus no
c	111	144.2	14.4	96396	2	AC133490 AC133490 Rattus no
c	112	144.2	14.4	226287	2	AC123086 AC123086 Rattus no
	113	143.2	14.3	189272	2	AC119503 AC119503 Rattus no
	114	142.8	14.3	236086	2	AC107243 AC107243 Rattus no
c	115	142.6	14.2	251625	2	AC095726 AC095726 Rattus no
	116	142.4	14.2	240800	2	AC095964 AC095964 Rattus no
	117	142.4	14.2	274697	2	AC098240 AC098240 Rattus no
	118	142	14.2	278905	2	AC126641 AC126641 Rattus no
	119	141.8	14.2	164138	2	AC109432 AC109432 Rattus no
c	120	141.8	14.2	264297	2	AC094040 AC094040 Rattus no
c	121	141.6	14.1	239057	2	AC115637 AC115637 Rattus no
c	122	141.6	14.1	280417	2	AC095158 AC095158 Rattus no
c	123	141.4	14.1	209093	2	AC109607 AC109607 Mus muscu
	124	141.2	14.1	190073	2	AC138388 AC138388 Mus muscu
	125	141.2	14.1	229887	2	AC073297 AC073297 Mus muscu
	126	140.6	14.0	110000	2	AC115960_0 AC115960 Mus muscu
	127	140.6	14.0	187446	2	AC124374 AC124374 Mus muscu
	128	140.6	14.0	210320	2	AC140324 AC140324 Mus muscu
	129	140.6	14.0	225027	2	AC131725 AC131725 Mus muscu
	130	140.4	14.0	245248	2	AC117869 AC117869 Rattus no
c	131	139.8	14.0	242619	2	AC115464 AC115464 Rattus no
	132	139	13.9	76861	2	AC087121 AC087121 Mus muscu
c	133	137.8	13.8	190175	2	AC137330 AC137330 Rattus no
	134	137.8	13.8	218800	2	AC094410 AC094410 Rattus no
	135	137.8	13.8	247754	2	AC125950 AC125950 Rattus no
	136	137.4	13.7	231108	2	AC141480 AC141480 Mus muscu
c	137	136.6	13.6	39723	9	AC141930 AC141930 Homo sapi
c	138	136.4	13.6	253951	2	AC128478 AC128478 Rattus no
c	139	136.2	13.6	213248	2	BX465859 BX465859 Danio rer
	140	136	13.6	212491	10	AL929228 AL929228 Mouse DNA
	141	136	13.6	230283	2	AC119802 AC119802 Mus muscu
c	142	136	13.6	246070	2	AC135737 AC135737 Mus muscu
	143	135.8	13.6	267991	2	AC112578 AC112578 Rattus no
c	144	135.6	13.5	192613	2	AC141326 AC141326 Rattus no
c	145	135.2	13.5	219108	2	AC095355 AC095355 Rattus no
c	146	134.8	13.5	243164	2	AC132173 AC132173 Rattus no
	147	134.8	13.5	253902	2	AC095908 AC095908 Rattus no
c	148	134.6	13.4	194204	2	AC131531 AC131531 Rattus no
	149	134.6	13.4	201679	2	AC133459 AC133459 Mus muscu
	150	134.6	13.4	212253	2	AC140184 AC140184 Mus muscu
c	151	134.6	13.4	219486	2	AC116283 AC116283 Rattus no
	152	134.6	13.4	222256	2	AC134482 AC134482 Rattus no
c	153	134.2	13.4	133719	14	AF083501 AF083501 Macaca mu
	154	134.2	13.4	160553	5	AL807792 AL807792 Zebrafish
	155	134	13.4	90406	2	AC133297 AC133297 Rattus no
c	156	133.6	13.3	212977	2	AC134108 AC134108 Rattus no
	157	133.2	13.3	231921	2	AC127906 AC127906 Rattus no
	158	133	13.3	230183	2	AC107559 AC107559 Rattus no
	159	132.8	13.3	98228	2	AC135405 AC135405 Rattus no
c	160	132.8	13.3	265401	2	AC099177 AC099177 Rattus no
	161	132.6	13.2	253951	2	AC128478 AC128478 Rattus no
c	162	132	13.2	313488	2	AC128784 AC128784 Rattus no
	163	131.8	13.2	261251	2	AC098618 AC098618 Rattus no
	164	131.6	13.1	208254	2	AC119002 AC119002 Rattus no
c	165	131.6	13.1	241649	2	AC120675 AC120675 Rattus no
	166	131.6	13.1	241726	2	AC097121 AC097121 Rattus no
c	167	131.4	13.1	265229	2	AC121468 AC121468 Rattus no
c	168	130.8	13.1	252686	2	AC106177 AC106177 Rattus no
c	169	130.4	13.0	110000	2	AC113631_0 AC113631 Rattus no
c	170	130.4	13.0	152589	2	BX510343 BX510343 Danio rer
c	171	130.2	13.0	205843	2	AC141957 AC141957 Rattus no
	172	130.2	13.0	250264	2	AC108970 AC108970 Rattus no
c	173	130	13.0	161984	9	AC106772 AC106772 Homo sapi
c	174	129.4	12.9	256759	2	AC126002 AC126002 Rattus no
	175	129.2	12.9	199495	2	AC079218 AC079218 Mus muscu
	176	128.6	12.8	196947	10	AC084826 AC084826 Genomic s
	177	128.6	12.8	228588	2	AC096241 AC096241 Rattus no
	178	128.4	12.8	181326	10	AL929165 AL929165 Mouse DNA
c	179	128.4	12.8	224052	2	AC124316 AC124316 Mus muscu
	180	128.4	12.8	238039	2	AC140796 AC140796 Mus muscu
c	181	128	12.8	132948	2	AC006410 AC006410 Homo sapi
c	182	127.6	12.7	35100	6	AR065849 AR065849 Sequence
c	183	127.6	12.7	35100	6	AR127847 AR127847 Sequence
c	184	127.6	12.7	35100	6	AR194749 AR194749 Sequence
c	185	127.6	12.7	137508	14	KSU75698 U75698 Kaposi's sa
	186	127.4	12.7	154918	9	AL138781 AL138781 Human DNA
c	187	127.4	12.7	163531	2	AL929518 AL929518 Danio rer
c	188	127.2	12.7	178273	2	AC005308 AC005308 Plasmodiu
	189	127.2	12.7	253132	3	AE014846 AE014846 Plasmodiu
c	190	127	12.7	199474	10	AC122873 AC122873 Mus muscu
c	191	126.8	12.7	230310	2	AC120079 AC120079 Rattus no
	192	126.8	12.7	293976	2	AC133832 AC133832 Rattus no
c	193	126.4	12.6	231698	2	AC118117 AC118117 Rattus no
	194	126.4	12.6	263700	2	AC098196 AC098196 Rattus no
c	195	125.8	12.6	77436	2	AC004581 AC004581 Homo sapi
	196	125.8	12.6	147854	9	AC021422 AC021422 Homo sapi
c	197	125.8	12.6	161644	9	AL161645 AL161645 Human DNA
c	198	125.8	12.6	179236	2	AC005143 AC005143 Homo sapi
c	199	125.8	12.6	288551	2	BX322534 BX322534 Homo sapi
c	200	125.6	12.5	111023	9	AL592464 AL592464 Human DNA
c	201	125.6	12.5	184104	9	AC020709 AC020709 Homo sapi
c	202	125.4	12.5	113891	5	AL713987 AL713987 Zebrafish
	203	125	12.5	220015	2	AC123418 AC123418 Rattus no
	204	125	12.5	247380	2	AC127159 AC127159 Rattus no
	205	124.8	12.5	224024	10	AL662802 AL662802 Mouse DNA
c	206	124.6	12.4	172485	2	AC141978 AC141978 Rattus no
c	207	124.6	12.4	181376	2	AC114990 AC114990 Mus muscu
c	208	124.6	12.4	220897	2	AL954635 AL954635 Homo sapi
c	209	124.6	12.4	238725	2	AC130214 AC130214 Mus muscu
	210	124.6	12.4	246075	2	AC111469 AC111469 Rattus no
	211	124.4	12.4	231864	2	AC097246 AC097246 Rattus no

212	123.8	12.4	146824	2	AC118745	AC118745 Mus muscu
c 213	123.8	12.4	164138	2	AC109432	AC109432 Rattus no
c 214	123.8	12.4	212899	2	AC102580	AC102580 Mus muscu
215	123.4	12.3	240525	2	AC096895	AC096895 Rattus no
c 216	123	12.3	284984	2	AC107121	AC107121 Rattus no
c 217	122.8	12.3	86165	5	AL954681	AL954681 Zebrafish
218	122.8	12.3	266712	2	BX088533	BX088533 Danio rer
c 219	122.6	12.2	179192	2	AC044848	AC044848 Mus muscu
220	122.6	12.2	220050	10	AL583887	AL583887 Mouse DNA
c 221	122.6	12.2	236344	10	AL583889	AL583889 Mouse DNA
c 222	122.4	12.2	112085	9	AC022132	AC022132 Homo sapi
c 223	122.4	12.2	112085	9	AC026712	AC026712 Homo sapi
c 224	122.4	12.2	139111	9	AC008768	AC008768 Homo sapi
c 225	122.4	12.2	183330	2	AC008159	AC008159 Homo sapi
c 226	122.4	12.2	198147	2	AC115032	AC115032 Mus muscu
c 227	122.4	12.2	201702	10	AL772335	AL772335 Mouse DNA
228	122.2	12.2	250286	2	AC119640	AC119640 Rattus no
229	122	12.2	41951	2	AC139177	AC139177 Ciona sav
230	122	12.2	292785	2	AC134591	AC134591 Mus muscu
231	122	12.2	321062	2	AC118593	AC118593 Mus muscu
232	121.8	12.2	204317	2	AC119659	AC119659 Rattus no
233	121.8	12.2	211758	2	AC095451	AC095451 Rattus no
c 234	121.6	12.1	201328	10	AL840635	AL840635 Mouse DNA
c 235	121.2	12.1	84563	5	AC096885	AC096885 Danio rer
236	120.8	12.1	168153	2	AC103388	AC103388 Mus muscu
237	120.6	12.0	241394	2	AC131549	AC131549 Rattus no
238	120.6	12.0	261236	2	AC119381	AC119381 Rattus no
c 239	120.4	12.0	190987	2	AC113883	AC113883 Rattus no
c 240	120.4	12.0	254372	2	AC132548	AC132548 Rattus no
c 241	120	12.0	174231	9	AC021054	AC021054 Homo sapi
242	120	12.0	180485	10	AL683823	AL683823 Mouse DNA
243	119.8	12.0	222259	2	AC110858	AC110858 Rattus no
244	119.8	12.0	261603	2	AC098318	AC098318 Rattus no
245	119.4	11.9	238653	2	AC099453	AC099453 Rattus no
246	119.2	11.9	110000	2	AC110828_3	Continuation (4 of
247	119.2	11.9	124636	9	AP000721	AP000721 Homo sapi
248	119.2	11.9	286029	2	AC110845	AC110845 Rattus no
249	119.2	11.9	310991	2	AC109011	AC109011 Rattus no
250	119	11.9	236394	2	AC120124	AC120124 Mus muscu
c 251	119	11.9	251506	2	AC106244	AC106244 Rattus no
c 252	118.8	11.9	131112	9	AL157712	AL157712 Human DNA
c 253	118.6	11.8	120084	9	AC006205	AC006205 Homo sapi
254	118.4	11.8	159117	2	AC132272	AC132272 Mus muscu
255	118.4	11.8	208049	2	AC108342	AC108342 Rattus no
c 256	118.4	11.8	223311	2	AC094165	AC094165 Rattus no
257	118.2	11.8	565	11	BV015874	BV015874 S208P6661
258	118.2	11.8	231877	2	AC121634	AC121634 Rattus no
c 259	118.2	11.8	239099	2	AC108535	AC108535 Rattus no
c 260	117.8	11.8	258510	2	AC130514	AC130514 Rattus no
c 261	117.6	11.7	185500	2	AC118940	AC118940 Mus muscu
262	117.6	11.7	188994	2	AC110325	AC110325 Rattus no
263	117.6	11.7	192453	2	AC023920	AC023920 Homo sapi
264	117.6	11.7	212419	2	AC099586	AC099586 Mus muscu
265	117.6	11.7	305997	2	AC123003	AC123003 Rattus no
266	117.4	11.7	68595	2	AC084772	AC084772 Homo sapi
c 267	117.4	11.7	193067	2	BX323458	BX323458 Danio rer
268	117.2	11.7	221568	2	AC103452	AC103452 Rattus no
c 269	117.2	11.7	252813	2	AC118528	AC118528 Rattus no
c 270	117	11.7	130364	9	AC116316	AC116316 Homo sapi
c 271	116.8	11.7	240279	2	AC126486	AC126486 Rattus no
c 272	116.8	11.7	270456	2	AC111575	AC111575 Rattus no
c 273	116.4	11.6	87233	2	AC141102	AC141102 Rattus no
274	116.2	11.6	197014	2	AC124133	AC124133 Mus muscu
c 275	116.2	11.6	201675	2	AC073795	AC073795 Mus muscu
276	116	11.6	150612	2	AC126099	AC126099 Rattus no
277	116	11.6	172533	2	AC120551	AC120551 Mus muscu
c 278	116	11.6	241822	2	AC127138	AC127138 Rattus no
c 279	115.8	11.6	172295	10	AL845460	AL845460 Mouse DNA
280	115.8	11.6	173641	2	AC141089	AC141089 Rattus no
281	115.8	11.6	183111	2	AC099708	AC099708 Mus muscu
c 282	115.8	11.6	184986	2	BX248111	BX248111 Danio rer
283	115.8	11.6	219092	2	AC136030	AC136030 Rattus no
284	115.8	11.6	258815	2	AC127817	AC127817 Rattus no
c 285	115.6	11.5	177962	9	AC008763	AC008763 Homo sapi
286	115.6	11.5	233962	2	AC094754	AC094754 Rattus no
287	115.6	11.5	243164	2	AC132173	AC132173 Rattus no
c 288	115.6	11.5	253902	2	AC095908	AC095908 Rattus no
c 289	115.6	11.5	276624	2	AC132560	AC132560 Rattus no
290	115.4	11.5	168958	10	AL591712	AL591712 Mouse DNA
291	115.4	11.5	175636	9	AC021744	AC021744 Homo sapi
292	115.4	11.5	183863	2	AC110455	AC110455 Rattus no
293	115.4	11.5	204019	2	AC139412	AC139412 Rattus no
c 294	115.4	11.5	235188	2	AC108344	AC108344 Rattus no
c 295	115.4	11.5	263387	2	AC112573	AC112573 Rattus no
296	115	11.5	233663	2	AC115486	AC115486 Rattus no
297	114.8	11.5	259487	2	AC124756	AC124756 Mus muscu
298	114.8	11.5	304681	2	AC016483	AC016483 Mus muscu
c 299	114.6	11.4	224199	2	AC110122	AC110122 Rattus no
300	114.6	11.4	239392	2	AC115417	AC115417 Rattus no

ALIGNMENTS

```
RESULT 1
AF083501
LOCUS       AF083501               133719 bp    DNA     linear   VRL 26-JUN-2000
DEFINITION  Macaca mulatta rhadinovirus 17577, complete genome.
ACCESSION   AF083501
VERSION     AF083501.3  GI:8714565
KEYWORDS    .
SOURCE      Macaca mulatta rhadinovirus 17577
  ORGANISM  Macaca mulatta rhadinovirus 17577
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Rhadinovirus.
REFERENCE   1  (bases 1 to 133719)
  AUTHORS   Searles,R.P., Bergquam,E.P., Axthelm,M.K. and Wong,S.W.
  TITLE     Sequence and genomic analysis of a Rhesus macaque rhadinovirus with
            similarity to Kaposi's sarcoma-associated herpesvirus/human
            herpesvirus 8
  JOURNAL   J. Virol. 73 (4), 3040-3053 (1999)
  MEDLINE   99174001
   PUBMED   10074154
REFERENCE   2  (bases 1 to 133719)
  AUTHORS   Searles,R.P., Bergquam,E.P., Axthelm,M.K. and Wong,S.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1998) Division of Pathobiology and Immunology,
            Oregon Health Sciences University/ Oregon Regional Primate Research
            Center, 505 NW I85th Avenue, Beaverton, OR 97006, USA
REFERENCE   3  (bases 1 to 133719)
  AUTHORS   Searles,R.P., Bergquam,E.P., Axthelm,M.K. and Wong,S.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-1999) Division of Pathobiology and Immunology,
            Oregon Health Sciences University/ Oregon Regional Primate Research
            Center, 505 NW I85th Avenue, Beaverton, OR 97006, USA
  REMARK    Sequence update by submitter
REFERENCE   4  (bases 1 to 133719)
  AUTHORS   Searles,R.P., Bergquam,E.P., Axthelm,M.K. and Wong,S.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Division of Pathobiology and Immunology,
            Oregon Health Sciences University/ Oregon Regional Primate Research
            Center, 505 NW I85th Avenue, Beaverton, OR 97006, USA
  REMARK    Sequence update by submitter
COMMENT     On Jun 26, 2000 this sequence version replaced gi:5508840.
FEATURES             Location/Qualifiers
     source          1. .133719
                     /organism="Macaca mulatta rhadinovirus 17577"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:83534"
     repeat_region   1. .840
                     /note="partial terminal repeat"
     CDS             1353. .2624
                     /codon_start=1
                     /product="R1"
                     /protein_id="AAD21330.1"
                     /db_xref="GI:4494908"
```

/translation="MFVLVLFMLLQPVSVELLPAKLTSVPTWCPPHPGDTYLLTCRGT
STARDQRSTQWFRNNTLMRGSNFYGRLVSVTPNATISDRYACQTKTTTRSNNIDFRVS
SSRLTLQERCSSYGYTYANNTRVLRCYSGGNVTLRNVVFHLNGTAVINGTTTNIHTFV
LTEKTGGTYFCSAFIGNEKFYSQTINVFFTSFTFKPTNDIPNESHFNKTGQIQQTASV
QHPENYVVFSVPVFSIGVLTGIAISLIMCWLFTIRCNENSESSTNSYASQTSYIQPSH
NQRSNTNECSRHTYRNAHQEESIEELPNQHTSETDSCCQLVLLEVKNVAYDGPQENTI
NEVMEQYDDVVVKNIEQTSYEDNVEHMDYSDTINPNFNYYSGLILEEVDEVFYNELEN
QYHGLILENLDHNEYNHLNELNMIEQYDWLE"

CDS complement(2692. .3258)
/note="ORF 2; similar to Kaposi's sarcoma-associated herpesvirus ORF 2"
/codon_start=1
/product="dihydrofolate reductase"
/protein_id="AAD21331.1"
/db_xref="GI:4494909"
/translation="MDITVNCIVAVDEQLGIGKNGTMPWPYLRNEMMYFQKMTSTPSV
VGEKNVVIMGKRTWFSIPEKKRPLVNRINIILSRELREPPHGAHFLARTLDDAFNFYR
QYKLKEQLNTVWVIGGKSVYESVLNYKCPLKLYITRIMESFDCDVFFPSINFTEYTML
SEIPGKDTNFEENGIKYKFQVYEKNFNK"

CDS 3676. .5613
/note="ORF 4; similar to Kaposi's sarcoma-associated herpesvirus ORF 4"
/codon_start=1
/product="complement binding protein"
/protein_id="AAD21332.1"
/db_xref="GI:4494910"
/translation="MTFKLFPLFLLHAIMYVHCDENCKPPHFTEYRVKSNTEKDLYSV
GETAELICRPGYVTNTKIITTECLQNGTWSTPNFPCDRKRCPTPADLLNGAVHIHGGD
NALKFGSNISYECNEGYDLIGSNVRFCILQDTENVNWDSNEPVCEIQKCIKPPAVEHG
DYLPNQDVYNYGDAITFKCSLSYTLVGSTTLVCTSNKKWSNSFPTCLMLVCESPQIDN
GYIDIGLSRRYNHGQSITVKCSDGYNIVGPETLTCTNTTWVPPLPKCVLVTNNPSTPM
PETPMPETPTPDYQKINLSTAKTATTPNAFVTTVVSPEKDDVTCVKPHFERFMVKAEN
DKEKYSVGASVELICRPGFTKMQSTVSVECLSNGTWTAPNAKCHRKKCPTPQELLNGE
YIVTSGEDAFKYGTNITYKCNEGYQLLGSMVRICMLKDDLKTVDWEPKAPICDIEKCK
PPPQITNGKYHPVKDFYQYLDTVTFSCNRDFSLVGDEMTTCISNTWNKPFPRCEQITC
SAPNIAHGKLLTGSSSVYKYGQSVTIGCETGFTLIGSEISTCKDSSWDPPLPTCVPAV
SMPSDTPKPETKKPNTPTPEAPKPNTPNVGTHTPFKPPPQNPPIAPPMSKWKRHVVLV
LFASVASLLFVLAALYCCFLK"

CDS 6045. .9443
/note="ORF 6; similar to Kaposi's sarcoma-associated herpesvirus ORF 6"
/codon_start=1
/product="ssDNA binding protein"
/protein_id="AAD21333.1"
/db_xref="GI:4494911"
/translation="MASKGNAGQPLEDNQGSRAPIGACGYVYAYSKQDFPFAEASILG
NRPSGSGVFSLPILYGLTVEHEFPLTVKAAYKKVDTTTLAVKVTCFHREVIVFHNASL
FRPVFDGTGLNELCEEARALFGYTQFIEPGPPHSIWNPLECPQLPDKDEMFLGVVVTE
GFKERLWRGCLVPAVFQTQQVQIAGRQAFKVPLYDEDLFAPHGHRMPRFYHKDVSAYL
YDSLFTSIAQALRLKDVTAVIHATEKQFMQDHYKIAKIVQAKQFSTTLPKTTDGSSHM
IVDSVVAELALSYGCMFLECPQDACELLNYDSWPIFDGCDSPEARVNALERWSAEQAV
HVAGQLFAANSVLYLTKVQKQAPRGQKGDVNVYNSFFLQHGLGFLNEATIKENGSEAF
KGVPSNALDGSSFTPYHLAYAASFSPHLLAKLCYYMQFLQHHKSSTNQAFNMVHYVGT
AANSEMCTLCHGNTPATCLNTLFYRLKDRFPAVTTPQRRDPYVVTGTAGTFNDLEILG
NFASFRDREEDGNPADEHPKYTYWQLCQTVTEKLSAIGITEDHDNHVNLITNIQSFLR
VFKGIDSIVDGEVMKFVNSMIKNNFNFREHVKSVHHILQFCCNVYWQAPCAVFLNLYY
KSLLWIIQDICLPYCMIYEQDNPAMGILPSEWLKMHFQTLWTNFKAACLDRGVLTGCE
LKIVHRDMFCDFFDTDAGSNGLMAPFKMQVRIARAMMVVPKSIKIKNRIIFSNTAGSE
AVQSGFVKPTGTRDTYVVAGPYMKFLNSLHRALFPDTKTAALYLWHKISQTNKTPVLK
DVPDDELAELVSYVKTNSLAFEETNVLDVVPDSLMSYARIKLNGAILRACGQIQFYAT
TLHCLTPVLQTIDAEEYPHVLGSAAIATPVAYLAEIRGRTALTVQTTARQPVAATGRL
RPVITVPMVVNKYTGVNGNNNVFHCGNLGYFAGRGVDRNLWPESSPFKKTGVSAMLRK
RHVMMTPIIDRLIKRAAGQTISTFEAESVKRSVQALLEDKDNPNLLKSVILELIRHLG
KGCQDLSSEDVQYYLGDYCMLTDEVLFTLDNIAQSGVPWTIEDAGALIEDRQDADDLQ
FVDSDDIATASCQPPEEQLPTPSAGALLAGKKRKINALLSDLDL"

CDS 9468. .11528
/note="ORF 7; similar to Kaposi's sarcoma-associated herpesvirus ORF 7"
/codon_start=1
/product="transport protein"
/protein_id="AAD21334.1"
/db_xref="GI:4494912"
/translation="MARELAALYAQLSALAVDLSLVIFADPRSIDGARILKTKTQIEN
LNRDLLPLLREQNSVETSSLSLEVEHLAKNIEDKLGELERSLRQRYSSREHFETLHLR
PECHYHSTVTFQFYGGGLIDVNMCLINDVELLCKRLGSVFYCIGANEALSGLNRVLTF
LSTLRGISPIPHPDLYVTSVPCVQCLREIELVPNQGSSLLAVLADRHCDHLCKKVRAE
PIHGLFETELSQLGLKVTKRSDATQHGVRSSADQLRESSLAAIQDHNIFKRVSASIME
LSNLIYWNAGQTGLQTGTENECSQMARLLTHEADMHEHRALITPKLSATHFYDCFRPD
PIESLFCGGLFNSIDDTINALSRDCSVTFFQQANYTNVMRKQNELFTRLNSILRQGSA
GSQKPATPSEPRTTTVAATAASDVIKDAQYRKEQYMKKVARDGFKKLTECLQTQSAVL
ANALCMRVWGGVAYGEASELVNHFLLRRRFVALPWEARCRSDQILFENSKYIKNSLYS
QRLSREHVEIITLQFYGLITGPLTRQSDLFPGPANVALAQCFEAAGMLPHHKMLVSEM
IWPQIQPKDWIDQTFNRFYQLPEGDLNAVQKSAWCFIRELVLSVALYNRTWEKTLRIF
SLAREKLSISNLDVKGLTSGLYLTYEQDAPLVLISQNTGWIFKDLYALLYHHLQLSDG
HDDN"

CDS 11515. .14004
/note="ORF 8; similar to Kaposi's sarcoma-associated herpesvirus ORF 8"
/codon_start=1
/product="glycoprotein B"
/protein_id="AAD21335.1"
/db_xref="GI:4494913"
/translation="MMITNRTRRLLRAWVVIIAIGTAVGENVTTPKGATTTAKPTPGP
STPTPPENPPRAEAFKFRVCSASATGELFRFNLEKTCPGTEDKTHQEGILMVFKKNIV
PHIFKVRRYRKVATSVTVYRGWTETAVTGKQEVIRPVPQYEINHMDTTYQCFSSMRVN
VNGIVNTYTDRDFTNQTVFLQPVEGLTDNIQRYFSQPVLYTTPGWFPGIYRVRTTVNC
EIVDMIARSAEPYSYFVTALGDTVEVSPFCHNDSTCSVAEKTENGLGARVLTNYTIVD
FATRQPTTETRVFADSGEYTVSWKAEDPKSAVCALTLWKTFPRAIQTTHEASYHFVAN
DVTATFTSPLSQVTNFTGTYPCLNDVIQKTLNATIKKLSDTHATNGSEQYYETEGGLF
LLWQPLTPLSLADEMRELNGTTPAPPTTTSTANRVRRSVGTNEQATDDLAAPQLQFAY
DKLRASINKVLEELSRAWCREQVRDTYMWYELSKINPTSVMTAIYGRPVSAKFVGDAI
SVTDCVAVDQASVSIHKSLRTSTPGICYSRPPVTFRFLNSTTLFKGQLGPRNEIILTD
NQVEACKETCEHYFIASNVTYYYKDYVFVKKINTSEISTLGTFIALNLSFIENIDFRV
IELYSRAEKKLSGSVFDIETMFREYNYYTQRLAGLREDLDNTIDLNRDRLARDLSEIV
ADLGDVGRTVVNVASSVITLFGSIVSGFINFIKSPFGGMLMILVIVAVVLIVFALNRR
TNAIAQAPIRMIYPDIDKMQPSGGKVDQEQIKNILAGMHQLQQEERRRLDEQQRSAPS
LFRRASDGLKRRFRGYKPLENEEAQEYEMSK"

CDS 14122. .17166
/note="ORF 9; similar to Kaposi's sarcoma-associated herpesvirus ORF 9"
/codon_start=1
/product="DNA polymerase"
/protein_id="AAD21336.1"
/db_xref="GI:4494914"
/translation="MDFFNPYLGPRGPRPHSHRGTDAPAPAGAGAVQPPPDVCRLIPA
CLRTPGAGGMIPVTIPFPPTYFENGARGDVLLANERSMWTARDRKPVAPDPQDQSITF
HAYDVVETTYAADRCAEVPSRFQTDIIPSGTVLKLLGRTEDGTSVCVNVFRQQVYFYA
KVPAGINVTHILQQALKNTAGRAACGFSTRRVNKRILKTYDVAEHPVTEITLSSGSML
STLSDRLVACGCEVFESNVDAVRRFVLDHGFTTFGWYSCARATPRLAARDARTALEFD
CSWEDLSVQADRSDWPPYRIVAFDIECTGEAGFPCATRDGDAVIQISCVFYTTREGAP
NPPNILFSVGTCDPIPDTDVLEFPSEYDMLVSFFAMIRDFEVDFLTGYNISNFDLPYL
ITRASQVYNLRLNEYTKIKTGSIFEVHEPRGGGGGFMRSVSKIKIAGIVPIDMYQVCR
EKLSLSDYKLDTVARQCLGGKKEDVSYKDIPPLFRSGPGGRAKVGSYCVMDSVLVMDL
LKMFMIHVEISEIAKLAKIQARRVLTDGQQLRVFSCLLEAAARENFILPVPTPEGQGG
YQGATVINPIPGFYDEPVLVVDFASLYPSIIQAHNLCYSTMIHGRDLHLHPNLTPDDY
ETFVLSGGPVHFVKKHKRESLLGRLLTVWLEKRRAIRRTLAACDDPSLKTILDKQQLA
IKVTCNAVYGFTGVASGLLPCINIAETVTLRGRTMLEMSKSYVEALTTEDLRTRLGRE
VTARHGARFRVVYGDTDSLFIACDGYSAEAVSAFCDDLAARITADLFPPPIKLEAEKT
FKCLLLLTKKRYIGVLLNDKMVMKGVDLIRKTACKFVQERCRAILDLVLHDPEVKAAA
RLLCKRPPHAVYEEGLPAGFIKIVEVLNASYLDLRNSVVPIEQLTFSTELSRPVCDYK
TTNLPHLAVYQKLASRCEELPQVHDRIPYVFVDAPGSLKSDLAEHPDYVRQHQIPVAV
DLYFDKLVHGAANILQCLFGNNADTTVAILYNFLNVPYKLFS"

CDS 17261. .18511
/note="ORF 10; similar to Kaposi's sarcoma-associated herpesvirus ORF 10"
/codon_start=1
/product="unknown"
/protein_id="AAD21337.1"
/db_xref="GI:4494915"
/translation="MLVNELSVVLGDWEVTFHRGRFSFVNLTRLQTFKGHGGYARVRL
PFSLDQLLHQHFAFGLVTRLKELPPFSDCVALIAPLDSGGDADAARVAPGFVLDSSRP
LTVWVNASGRHTIRFCLLFLKPIDLERAVTYVFGENGGARSEGTPKPTCATESLPGGP
LRVSGEASQTSPHSFVAYFPTANSVACLSLLRLQVRPFSDDAAHRDARISPKYVTFSN
SGGNVCKASVHTLSPSRCKTAQMEIIYAPGDPNAEIVLGQSGPVLPTHTGGRVLGVYA
DAEKTIQPGSSAEVRVQLIFQQGAAARGDLAFLVTGVAPEPLFVVTPALLLSGCTTHL
RLFNPNGTPTTIKRDTLVAAAAPCPVVRLSSADDAPRDLVASPDTGALSINAFTIPVG

```
                FPGVVSAECHVSLRDNGVHERMNH"
     CDS        18520. .19749
                /note="ORF 11; similar to Kaposi's sarcoma-associated
                herpesvirus ORF 11"

  Query Match            100.0%;  Score 1001;  DB 14;  Length 133719;
  Best Local Similarity  100.0%;  Pred. No. 4e-150;
  Matches 1001;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CGCCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     132719 CGCCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTG 132778

Qy         61 GTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     132779 GTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGG 132838

Qy        121 GGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     132839 GGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGAC 132898

Qy        181 AGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     132899 AGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGC 132958

Qy        241 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     132959 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 133018

Qy        301 AGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133019 AGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 133078

Qy        361 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133079 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 133138

Qy        421 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133139 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 133198

Qy        481 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133199 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 133258

Qy        541 CTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133259 CTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGA 133318

Qy        601 TGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133319 TGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 133378

Qy        661 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133379 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 133438

Qy        721 TGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133439 TGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 133498

Qy        781 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133499 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGC 133558

Qy        841 TTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133559 TTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAG 133618

Qy        901 GGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACCCGGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133619 GGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACCCGGG 133678
```

```
Qy         961 CCTCTCCCCCGCCCGGACCGCCGGGCACCCGGCCCAGATCT 1001
               |||||||||||||||||||||||||||||||||||||||||
Db      133679 CCTCTCCCCCGCCCGGACCGCCGGGCACCCGGCCCAGATCT 133719
```

RESULT 2
AC116487/c
LOCUS AC116487 148106 bp DNA linear HTG 12-JUL-2002
DEFINITION Mus musculus clone RP23-428I9, WORKING DRAFT SEQUENCE, 22 ordered pieces.
ACCESSION AC116487
VERSION AC116487.2 GI:21735111
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 148106)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP23-428I9
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 148106)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 148106)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (12-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Jul 12, 2002 this sequence version replaced gi:19774593.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

```
            ---------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ---------- Project Information
                Center project name: L23947
                Center clone name: 428_I_9
            ---------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 138537 bases at least Q40
                Consensus quality: 143163 bases at least Q30
                Consensus quality: 144845 bases at least Q20
                Insert size: 160000; agarose-fp
                Insert size: 146006; sum-of-contigs
                Quality coverage: 5.6 in Q20 bases; agarose-fp
                Quality coverage: 6.1 in Q20 bases; sum-of-contigs
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 22 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    13805: contig of 13805 bp in length
            *    13806    13905: gap of 100 bp
            *    13906    14560: contig of 655 bp in length
            *    14561    14660: gap of 100 bp
            *    14661    15529: contig of 869 bp in length
            *    15530    15629: gap of 100 bp
            *    15630    16744: contig of 1115 bp in length
            *    16745    16844: gap of 100 bp
            *    16845    18432: contig of 1588 bp in length
            *    18433    18532: gap of 100 bp
            *    18533    19971: contig of 1439 bp in length
            *    19972    20071: gap of 100 bp
            *    20072    21780: contig of 1709 bp in length
            *    21781    21880: gap of 100 bp
            *    21881    23204: contig of 1324 bp in length
            *    23205    23304: gap of 100 bp
            *    23305    24685: contig of 1381 bp in length
            *    24686    24785: gap of 100 bp
            *    24786    28365: contig of 3580 bp in length
            *    28366    28465: gap of 100 bp
            *    28466    30918: contig of 2453 bp in length
            *    30919    31018: gap of 100 bp
            *    31019    34321: contig of 3303 bp in length
            *    34322    34421: gap of 100 bp
            *    34422    36281: contig of 1860 bp in length
            *    36282    36381: gap of 100 bp
            *    36382    39099: contig of 2718 bp in length
            *    39100    39199: gap of 100 bp
            *    39200    42684: contig of 3485 bp in length
            *    42685    42784: gap of 100 bp
            *    42785    46395: contig of 3611 bp in length
            *    46396    46495: gap of 100 bp
            *    46496    51682: contig of 5187 bp in length
            *    51683    51782: gap of 100 bp
            *    51783    56221: contig of 4439 bp in length
            *    56222    56321: gap of 100 bp
            *    56322    66418: contig of 10097 bp in length
            *    66419    66518: gap of 100 bp
            *    66519    83479: contig of 16961 bp in length
            *    83480    83579: gap of 100 bp
            *    83580   110083: contig of 26504 bp in length
            *   110084   110183: gap of 100 bp
            *   110184   148106: contig of 37923 bp in length.
FEATURES            Location/Qualifiers
     source          1. .148106
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-428I9"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .13805
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    13906. .14560
                     /note="assembly_fragment"
     misc_feature    14661. .15529
                     /note="assembly_fragment"
     misc_feature    15630. .16744
                     /note="assembly_fragment"
     misc_feature    16845. .18432
                     /note="assembly_fragment"
     misc_feature    18533. .19971
                     /note="assembly_fragment"
     misc_feature    20072. .21780
                     /note="assembly_fragment"
     misc_feature    21881. .23204
                     /note="assembly_fragment"
     misc_feature    23305. .24685
                     /note="assembly_fragment"
     misc_feature    24786. .28365
                     /note="assembly_fragment"
     misc_feature    28466. .30918
                     /note="assembly_fragment"
     misc_feature    31019. .34321
                     /note="assembly_fragment"
     misc_feature    34422. .36281
                     /note="assembly_fragment"
     misc_feature    36382. .39099
                     /note="assembly_fragment"
     misc_feature    39200. .42684
                     /note="assembly_fragment"
     misc_feature    42785. .46395
                     /note="assembly_fragment"
     misc_feature    46496. .51682
                     /note="assembly_fragment"
     misc_feature    51783. .56221
                     /note="assembly_fragment"
     misc_feature    56322. .66418
                     /note="assembly_fragment"
     misc_feature    66519. .83479
                     /note="assembly_fragment"
     misc_feature    83580. .110083
                     /note="assembly_fragment"
     misc_feature    110184. .148106
                     /note="assembly_fragment"
BASE COUNT    36315 a  37882 c  38018 g  33788 t   2103 others
ORIGIN

  Query Match           28.2%;  Score 282.6;  DB 2;  Length 148106;
  Best Local Similarity   64.4%;  Pred. No. 3.6e-36;
  Matches  423;  Conservative    0;  Mismatches  234;  Indels    0;  Gaps    0;

Qy        191 GTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTG 250
              || ||  ||  |||  |       ||  |  ||  ||| |||| ||| ||||||||||
Db      40784 GTCTGTCTGTCTGTCTGTTTGTTTGTTTGTTTGTTTGCTTGCTTGCTTGCCTGCTTGCTT 40725

Qy        251 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 310
              | | |   || |||||||  ||| || |||||||      |   |   ||||||||
Db      40724 GCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCTT 40665

Qy        311 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 370
              |||     | | |   || |||||||  ||  ||||||||||     |   |   | ||
Db      40664 GCTTGCTTGCCTGCTTGCTTGCTTGCCTGCCTGCCTGCTTGCTTGCTTGCTTGCTTGCCT 40605

Qy        371 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 430
              ||| ||||||  |   | ||||| ||| || ||||||||||  ||  || ||||||| |
```

```
Db      40604 GCTTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCCTGCCTGCTTGCTTGCCT 40545

Qy        431 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              ||| ||| |||||| |||||||| ||| || ||| |||||||| ||| ||||||||||||
Db      40544 GCTTGCTTGCCTGCCTGCTTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTT 40485

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              ||| ||| || |||||||||||| ||| |||||||||||||| ||  || ||||||||
Db      40484 GCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTT 40425

Qy        551 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 610
                  |   |   | ||||||||||||||     |   |   |||||||| | ||||| |
Db      40424 GCCTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCCTGCCTGCTTGCTTGCTT 40365

Qy        611 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 670
              |   |  |||| |  |||| |   |  |||| |  |||| |   | ||||| |  |||
Db      40364 GCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGCCTGCTTGCTTGCCT 40305

Qy        671 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 730
              |   |  |||| | ||||| |   | || |  |  |||| |   |  |||| |  ||||
Db      40304 GCTTGCTTGCTTGCCTGCTTGCTTGCCTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 40245

Qy        731 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 790
              |   | || || |  |||| |   | |||||  |  |||| |   |  |||| |  ||||
Db      40244 GCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCTT 40185

Qy        791 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAG 847
              |   |  |||| | |||||  |   |  |||||||| |||||||||||||| |||| |
Db      40184 GCTTGCTTGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCTTGCTTGCTTGCCTGCTTG 40128
```

RESULT 3
AC099177
LOCUS AC099177 265401 bp DNA linear HTG 10-MAY-2003
DEFINITION Rattus norvegicus clone CH230-112E16, *** SEQUENCING IN PROGRESS ***, 2 unordered pieces.
ACCESSION AC099177
VERSION AC099177.6 GI:30522052
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 265401)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 265401)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 265401)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 10, 2003 this sequence version replaced gi:25139860.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
-------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
Center project name: GHAA
Center clone name: CH230-112E16
-------------- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 219762 bases at least Q40
Consensus quality: 223776 bases at least Q30
Consensus quality: 226617 bases at least Q20
Estimated insert size: 230040; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```
          * be preserved.
          *        1   263276: contig of 263276 bp in length
          *   263277   263376: gap of unknown length
          *   263377   265401: contig of 2025 bp in length.
FEATURES             Location/Qualifiers
     source          1. .265401
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-112E16"
     misc_feature    1. .1265
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    12027. .12919
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH365978"
     misc_feature    90541. .93662
                     /note="wgs_contig"
     misc_feature    136482. .137827
                     /note="wgs_contig"
     misc_feature    261732. .263276
                     /note="wgs_contig"
BASE COUNT    64515 a  48596 c  48625 g  66669 t  36996 others
ORIGIN

  Query Match             26.2%;  Score 262.2;  DB 2;  Length 265401;
  Best Local Similarity   61.5%;  Pred. No. 5.4e-33;
  Matches  420;  Conservative    0;  Mismatches  263;  Indels    0;  Gaps    0;

Qy        194 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 253
              ||| ||| ||   | |  |  |   | |||||||| ||||||||.|||||| ||||
Db     139005 TGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCTCGCTTGCCTGCCTGCTTGCT 139064

Qy        254 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 313
              | | |  ||||||||||  ||  ||||||||||      | | |   |||||| ||||
Db     139065 TGCCTGCCTGCCTGCTTGCCTGCCTGCCTGCTTGCCTGCCTGCCTGCCTGCCTGCT 139124

Qy        314 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 373
                  |   |   || ||| |||| ||| || ||| |||      |   |   | ||||
Db     139125 TGCCTGCTTGCCTGCTTGCCTGCTTGCTCGCTTGCCTGCCTGCTTGCTTGCCTGCCTGCC 139184

Qy        374 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 433
              ||||||  |   |    || ||  || |||||| |||  ||  ||||||||||| |||
Db     139185 TGCTTGCCTGCTTGCTCACTTGCCTGCTTGCTTGCCTGCCTGCCTGCCTGCTTGCCTGCC 139244

Qy        434 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 493
              ||| || || |  |||||  ||| ||||||||||||||| ||| || ||| ||| |||
Db     139245 TGCTTGCTTGTTCACTTGCCTGCTTGCCTGCTTGCTTGCTTGCTCGCTTGCCTGCCTGCC 139304

Qy        494 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGG 553
              ||| |||||| ||| |||  ||  ||||||  |||||||||||| ||||||  |||
Db     139305 TGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCTCGCTTGCCTGCCTGCCTGCT 139364

Qy        554 GGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTA 613
              | | |   | |||| ||| |||       | | |   |||||||| | ||| | | |
Db     139365 TGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCTCGCTTGCC 139424

Qy        614 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 673
              | ||||  | ||||  |   | ||||  | |||||  |   | |||||  |   ||| |
Db     139425 TGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCTTGCCTGCTTGCTCGCTTGCC 139484

Qy        674 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 733
              |  |||| |  |||  |   | || || | |||||  |      |||  |  |||| |
Db     139485 TGCTTGCTTGCTTGCCTGCCTGCCTGCTTGCCTGCTTGCTCACTTGCCTGCTTGCTTGCC 139544

Qy        734 GGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 793
              | || |  | |||||  |   |  ||||  |  || | |   | |||||  |  |||| |
Db     139545 TGCCTGCCTGCCTGCTTGCCTGCTTGCTTGCTTGTTCGCTTGCCTGCTTGCTTGCTTGCC 139604

Qy        794 GGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTAGTGGGGC 853
              | ||||  | ||||  |   | ||||||||  || |||||| ||||||||  |   | |

Db     139605 TGCCTGCCTGCCTGCCTGCTTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCTCGCTTGCC 139664

Qy        854 CGCTTGCCTGCTACTAGGGCTGC 876
                || ||||||||     | ||||
Db     139665 TGCCTGCCTGCTTGCCTGCCTGC 139687


RESULT 4
AC119239
LOCUS       AC119239               165180 bp    DNA     linear   HTG 10-MAR-2003
DEFINITION  Mus musculus clone RP24-228F21, WORKING DRAFT SEQUENCE, 6 unordered
            pieces.
ACCESSION   AC119239
VERSION     AC119239.3  GI:28894656
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 165180)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-228F21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165180)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 165180)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
```

```
COMMENT     On Mar 10, 2003 this sequence version replaced gi:28269570.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L25264
                Center clone name: 228_F_21
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 164030 bases at least Q40
                Consensus quality: 164471 bases at least Q30
                Consensus quality: 164601 bases at least Q20
                Insert size: 166000; agarose-fp
                Insert size: 164680; sum-of-contigs
                Quality coverage: 11.1 in Q20 bases; agarose-fp
                Quality coverage: 11.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1027: contig of 1027 bp in length
            *     1028     1127: gap of 100 bp
            *     1128    16448: contig of 15321 bp in length
            *    16449    16548: gap of 100 bp
            *    16549    76346: contig of 59798 bp in length
            *    76347    76446: gap of 100 bp
            *    76447   103034: contig of 26588 bp in length
            *   103035   103134: gap of 100 bp
            *   103135   153662: contig of 50528 bp in length
            *   153663   153762: gap of 100 bp
            *   153763   165180: contig of 11418 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165180
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-228F21"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .1027
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    1128. .16448
                     /note="assembly_fragment"
     misc_feature    16549. .76346
                     /note="assembly_fragment"
     misc_feature    76447. .103034
                     /note="assembly_fragment"
     misc_feature    103135. .153662
                     /note="assembly_fragment"
     misc_feature    153763. .165180
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    44313 a  40435 c  37850 g  42082 t    500 others
ORIGIN

  Query Match            25.7%;  Score 257.6;  DB 2;  Length 165180;
  Best Local Similarity  61.5%;  Pred. No. 3.3e-32;
  Matches  413;  Conservative    0;  Mismatches  259;  Indels    0;  Gaps    0;

Qy        194 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 253
              ||| ||||||    |     |  |    |  ||| ||| |||| | |  | ||||||||  |
Db     114816 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGGTTACTTGCTTGCTTGGT 114875

Qy        254 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 313
              |      || |||||||| ||| || ||||||||      |    |   || ||||||||
Db     114876 TGGTTACTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 114935

Qy        314 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 373
                |   |    | |||||||| ||| || ||||||||     |   |   |  ||||
Db     114936 TGCTTGCTTGGTTACTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 114995

Qy        374 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 433
              |||||||  |   |  |||| ||| || |||||| || || ||| |  | ||||||||||
Db     114996 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGGTTACTTGCTTGGTTACTTGCTTGCT 115055

Qy        434 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 493
              ||| || |||||||||||| ||| || |||||||||||| ||| || |||||| || ||
Db     115056 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGGTTACT 115115

Qy        494 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGG 553
              ||| || |||||||||||| ||| || |||||| || || ||| || ||||||||
Db     115116 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGGTTACTTGCTTGCTTGCTTGCTTGCT 115175

Qy        554 GGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTA 613
              |   |   | ||||||||||||     |   |    | ||| || | |||| | |
Db     115176 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGGTTACTTGCTTGCTTGCTTGCTTGCT 115235

Qy        614 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 673
              |  |||| |  |||| |   |  || | |  || | |   |  |||| |  |||| |
Db     115236 TGCTTGCTTGCTTGCTTGCTTGCTTGTTTGCTTGATTGATTGATTGCTTGCTTGCTTGCT 115295

Qy        674 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 733
              |  |||| |  |||| |   |  | || |  |||| |   |  |||| |  |||| |
Db     115296 TGCGTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 115355

Qy        734 GGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 793
              || | || |  |||| |   |  |||| |  |||| |   |  |||| |  |||| |
Db     115356 TGGTTACTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 115415

Qy        794 GGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGC 853
              |  |  |    |||| |   |  ||| |||| ||||||||||||||||||||  |   |
Db     115416 TGCTTAGTTACTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 115475

Qy        854 CGCTTGCCTGCT 865
              |||||| ||||
Db     115476 TGCTTGCTTGCT 115487


RESULT 5
AF210726
LOCUS       AF210726              130733 bp    DNA     linear   VRL 27-MAR-2000
DEFINITION  Macaca mulatta rhadinovirus 26-95 long unique region L-DNA,
            complete sequence.
ACCESSION   AF210726
VERSION     AF210726.1  GI:7329990
KEYWORDS    .
SOURCE      Macaca mulatta rhadinovirus.26-95
  ORGANISM  Macaca mulatta rhadinovirus 26-95
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Rhadinovirus.
REFERENCE   1  (bases 1 to 130733)
  AUTHORS   Alexander,L., Denekamp,L., Knapp,A., Auerbach,M.R., Damania,B. and
            Desrosiers,R.C.
  TITLE     The primary sequence of rhesus monkey rhadinovirus isolate 26-95:
            sequence similarities to Kaposi's sarcoma-associated herpesvirus
            and rhesus monkey rhadinovirus isolate 17577
  JOURNAL   J. Virol. 74 (7), 3388-3398 (2000)
  MEDLINE   20173730
   PUBMED   10708456
REFERENCE   2  (bases 1 to 130733)
  AUTHORS   Alexander,L., Denekamp,L.M., Knapp,A., Auerbach,M., Czajak,S.,
            Damania,B. and Desrosiers,R.C.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (02-DEC-1999) Microbiology, New England Regional Primate
            Research Center, One Pinehill Dr, Southborough, MA 01772, USA
FEATURES             Location/Qualifiers
     source          1. .130733
                     /organism="Macaca mulatta rhadinovirus 26-95"
                     /mol_type="genomic DNA"
                     /isolate="Macaca mulatta rhadinovirus isolate 26-95"
                     /db_xref="taxon:119193"
     gene            513. .1784
                     /gene="R1"
     CDS             513. .1784
                     /gene="R1"
                     /codon_start=1
                     /product="ORFR1"
                     /protein_id="AAF59980.1"
                     /db_xref="GI:7329991"
                     /translation="MFVLVLFMLLQPVSVELLPAKLTSVPTWCPPHPGDTYLLTCRGT
                     STARDQRSTQWFRNNTLMRGSNFYGRLVSVTPNATISDRYACQTKTTTRSNNIDFRVS
                     SSRLTLQERCSSYGYTYANNTRVLRCYSGGNVTLRNVVFHLNGTAVINGTTRNIHTFV
                     LTEKTGGTYFCSAFIGNEKFYSQTINVFFTSFTFKPTNDIPNESHFNKTGQIQQTASV
                     QHPENYVVFSVPVFSIGVLTGIAISLIMCWLFTIRCNENSESSTNSYASQTSYIQPSH
                     NQRSNTNECSRHTYRNAHQEESIEELPNQHTSETDSCCQLVLLEVKNVAYDGPQENTI
                     NEVMEQYDDVVVENIEQTSYEDNVEHMDYSDTINPNFNYYSGLILEEVDEVFYNELEN
                     QYHGLILENLDHNEYNHLNELNMIEQYDWLE"
     gene            complement(1852. .2418)
                     /gene="ORF02"
     CDS             complement(1852. .2418)
                     /gene="ORF02"
                     /codon_start=1
                     /product="DHFR"
                     /protein_id="AAF59981.1"
                     /db_xref="GI:7329992"
                     /translation="MDIAVNCIVAVDEQLGIGKNGTMPWPYLRNEMMYFQKMTSTPSV
                     VGEKNVVIMGKRTWFSIPEKKRPLVNRINIILSRELREPPHGAHFLARTLDDAFNFYR
                     QYKLKEQLNTVWVIGGKSVYESVLNYKCPLKLYITRIMESFDCDVFFPSINFTEYTML
                     SELPGKDTNFEENGIKYKFQVYEKNFNK"
     gene            2595. .3782
                     /gene="ORF04"
     CDS             2595. .3782
                     /gene="ORF04"
                     /codon_start=1
                     /product="complement binding protein"
                     /protein_id="AAF59982.1"
                     /db_xref="GI:7329993"
                     /translation="MFWIPITLCFICQVILVDSKDGENVTCLKPPFDRFMVKTANQNE
                     NYAVGTRVELICRPGFYKLQANVYVECLSNGTWTTPNAECRRKRCSNPEDILNGEVII
                     TDSDNAFKFGSNITYKCNTGYLLLGATVRTCLLKYDSNLVDWQPAAPTCEIEKCKKQP
                     DIENGKYYPVQEFYNYLETITFTCNKDFSLIGNTTTTCMTNGTWSSPVPKCQQITCSA
                     PNIDHGTLLVGSSRVYKHGQSVTIGCEDGFTLNGHKMCTCEYSLWNPPLPTCVPINKT
                     DPTPSEVPSPGTKEQEMPTPENPKSHESETTTTETPKTQTHKSETPSPKVPNPETHKP
                     TTPEAGISKQTTTNRPSKAPSQNPPIAPPMSKWKRHVVLVLFASVASLLFVLAALYCC
                     FLK"
     gene            4213. .7611
                     /gene="ORF06"
     CDS             4213. .7611
                     /gene="ORF06"
                     /codon_start=1
                     /product="ssDBP"
                     /protein_id="AAF59983.1"
                     /db_xref="GI:7329994"
                     /translation="MASKGNAGQPLEDNQGSRAPIGACGYVYAYSKQDFPFAEASILG
                     NRPSGSGVFSLPILYGLTVEHEFPLTVKAAYKKVDTTTLAVKVTCFHREVIVFHNASL
                     FRPVFDGTGLNELCEEARALFGYTQFIEPGPPHSIWNPLECPQLPDKDEMFLGVVVTE
                     GFKERLWRGCLVPAVFQTQQVQIAGRQAFKVPLYDEDLFAPHGHRMPRFYHKDVSAYL
                     YDSLFTSIAQALRLKDVTAVIHATEKQFMQDHYKIAKIVQAKQFSTTLPKTTDGSSHM
                     IVDSVVAELALSYGCMFLECPQDACELLNYDSWPIFDGCDSPEARVNALERWSAEQAV
                     HVAGQLFAANSVLYLTKVQKQAPRGQKGDVNVYNSFFLQHGLGFLNEATIKENGSEAF
                     KGVPSNALDGSSFTPYHLAYAASFSPHLLAKLCYYMQFLQHHKSSTNQAFNMVHYVGT
                     AANSEMCTLCHGNTPATCLNTLFYRLKDRFPAVTTPQRRDPYVVTGTAGTFNDLEILG
                     NFASFRDREEDGNPADEHPKYTYWQLCQTVTEKLSAIGITEDHDNHVNLITNIQSFLR
                     VFKGIDSIVDGEVMKFVNSMIKNNFNFREHVKSVHHILQFCCNVYWQAPCAVFLNLYY
                     KSLLWIIQDICLPYCMIYEQDNPAMGILPSEWLKMHFQTLWTNFKAACLDRGVLTGCE
                     LKIVHRDMFCDFFDTDAGSNGLMAPFKMQVRIARAMMVVPKSIKIKNRIIFSNTAGSE
                     AVQSGFVKPTGTRDTYVVAGPYMKFLNSLHRALFPDTKTAALYLWHKISQTNKTPVLK
                     DVPDDELAELVSYVKTNSLAFEETNVLDVVPDSLMSYARIKLNGAILRACGQIQFYAT
                     TLHCLTPVLQTIDAEEYPHVLGSAAIATPVAYLAEIRGRTALTVQTTARQPVAATGRL
                     RPVITVPMVVNKYTGVNGNNNVFHCGNLGYFAGRGVDRNLWPESSPFKKTGVSAMLRK
                     RHVMMTPIIDRLIKRAAGQTISTFEAESVKRSVQALLEDKDNPNLLKSVILELIRHLG
                     KGCQDLSSEDVQYYLGDYCMLTDEVLFTLDNIAQSGVPWTIEDAGALIEDRQDADDLQ
                     FVDSDDIATASCQPPEEQLPTPSAGALLAGKKRKINALLSDLDL"
     gene            7636. .9696
                     /gene="ORF07"
     CDS             7636. .9696
                     /gene="ORF07"
                     /codon_start=1
                     /product="transport protein"
                     /protein_id="AAF59984.1"
                     /db_xref="GI:7329995"
                     /translation="MARELAALYAQLSALAVDLSLVIFADPRSIDGARILKTKTQIEN
                     LNRDLLPLLREQNSVETSSLSLEVEHLAKNIEDKLGELERSLRQRYSSREHFETLHLR
                     PECHYHSTVTFQFYGGGLIDVNMCLINDVELLCKRLGSVFYCIGANEALSGLDRVLTF
                     LSTLRGISPIPHPDLYVTSVPCVQCLREIELVPNQGSSLLAVLADRHCDHLCKKVRAE
                     PIHGLFETELSQLGLKVTKRSDATQHGVRSSADQLRESSLAAIQDHNIFKRVSASIME
                     LSNLIYWNAGQTGLQTGTENECSQMARLLTHEADMHEHRALITPKLSATHFYDCFRPD
                     PIESLFCGGLFNSIDDTINALSRDCSVTFFQQANYTNVMRKQNELFTRLNSILRQGSA
                     GSQKPATPSEPRTTTVAATAASDVIKDAQYRKEQYMKKVARDGFKKLTECLQTQGAVL
                     ANALCMRVWGGVAYGEASELVNHFLLRRRFVALPWEARCRSDQILFENSKYIKNSLYS
                     QRLSREHVEIITLQFYGLITGPLTRQSDLFPGPANVVLAQCFEAAGMLPHHKMLVSEM
                     IWPQIQPKDWIDQTFNRFYQLPEGDLNAVQKSAWCFIRELVLSVALYNRTWEKTLRIF
                     SLAREKLSISNLDVKGLTSGLYLTYEQDAPLVLISQNTGWIFKDLYALLYHHLQLSDG
                     HDDN"
     gene            9683. .12172
                     /gene="ORF08"
     CDS             9683. .12172
                     /gene="ORF08"
                     /codon_start=1
                     /product="ORF08"
                     /protein_id="AAF59985.1"
                     /db_xref="GI:7329996"
                     /translation="MMITNRTRRLLRAWVVIIAIGTAVGENVTTPKGATTTAKPTPGP
                     STPTPPENPPRAEAFKFRVCSASATGELFRFNLEKTCPGTEDKTHQEGILMVFKKNIV
                     PHIFKVRRYRKVATSVTVYRGWTETAVTGKQEVIRPVPQYEINHMDTTYQCFSSMRVN
                     VNGIVNTYTDRDFTNQTVFLQPVEGLTDNIQRYFSQPVLYTTPGWFPGIYRVRTTVNC
                     EIVDMIARSAEPYSYFVTALGDTVEVSPFCHNDSTCSVAEKTENGLGVRALTNYTIVD
                     FATRKPTTETRVFADSGEYTVSWKAEDPKSAVCALTLWKTFPRAIQTTHEASYHFVAN
                     DVTATFTSPLSEVANFTGTYSCLDEVIQKTLNDTIKKLSDTHVTNGSAQYYKTEGGLF
                     LLWQPLTPLSLVDEMRGLNGTTPAPPATTSTVSRVRRSVNTNEQATDNLAAPQLQFAY
                     DKLRASINKVLEELSRAWCREQVRDTYMWYELSKINPTSVMTAIYGRPVSAKFVGDAI
                     SVTDCVAVDQASVSIHKSLRTSTPGMCYSRPPVTFRFLNSTTLFKGQLGPRNEIILTD
                     NQVEACKETCEHYFIASNVTYYYKDYVFVKKINTSEISTLGTFIALNLSFIENIDFRV
                     IELYSRAEKKLSGSVFDIETMFREYNYYTQRLAGLREDLDNTIDLNRDRLARDLSEIV
                     ADLGDVGRTVVNVASSVITLFGSIVSGFINFIKSPFGGMLMILVIVAVVLIVFALNRR
                     TNAIAQAPIRMIYPDIDKMQPSGGKVDQEQIKNILAGMHQLQQEERRRLDEQQRSAPS
                     LFRRASDGLKRRFRGYKPLENEEAQEYEMSK"
     gene            12290. .15334
                     /gene="ORF09"
     CDS             12290. .15334
                     /gene="ORF09"
                     /codon_start=1
                     /product="polymerase"
                     /protein_id="AAF59986.1"
                     /db_xref="GI:7329997"
                     /translation="MDFFNPYLGPRGPRPHSHRGTDAPAPAGAGAVQPPPDVCRLIPA
                     CLRTPGAGGMIPVTIPFPPTYFENGARGDVLLANERSMWTARDRKPVAPDPQDQSITF
                     HAYDVVETTYAADRCAEVPSRFQTDIIPSGTVLKLLGRTEDGTSVCVNVFRQQVYFYA
                     KVPAGINVTHILQQALKNTAGRAACGFSTRRVNKRILKTYDVAEHPVTEITLSSGSML
                     STLSDRLVACGCEVFESNVDAVRRFVLDHGFTTFGWYSCARATPRLAARDARTALEFD
                     CSWEDLSVQADRSDWPPYRIVAFDIECTGEAGFPCATRDGDAVIQISCVFYTTREGAP
                     NPPNILFSVGTCDPIPDTDVLEFPSEYDMLVSFFAMIRDFEVDFLTGYNISNFDLPYL
                     ITRASQVYNLRLNEYTKIKTGSIFEVHEPRGGGGGFMRSVSKIKIAGIVPIDMYQVCR
                     EKLSLSDYKLDTVARQCLGGKKEDVSYKDIPPLFRSGPGGRAKVGSYCVMDSVLVMDL
                     LKMFMIHVEISEIAKLAKIQARRVLTDGQQLRVFSCLLEAAARENFILPVPTPEGQGG
                     YQGATVINPIPGFYDEPVLVVDFASLYPSIIQAHNLCYSTMIHGRDLHLPHNLTPDDY
                     ETFVLSGGPVHFVKKHKRESLLGRLLTVWLEKRRAIRRTLAACDDPSLKTILDKQQLA
                     IKVTCNAVYGFTGVASGLLPCINIAETVTLRGRTMLEMSKSYVEALTTEDLRTRLGRE
```

```
                VTARHGARFRVVYGDTDSLFIACDGYSAEAVSAFCDDLAARITADLFPPPIKLEAEKT
                FKCLLLLTKKRYIGVLLNDKMVMKGVDLIRKTACKFVQERCRAILDLVLHDPEVKAAA
                RLLCKRPPHAVYEEGLPAGFIKIVEVLNASYLDLRNSVVPIEQLTFSTELSRPVCDYK
                TTNLPHLAVYQKLASRCEELPQVHDRIPYVFVDAPGSLKSDLAEHPDYVRQHQIPVAV
                DLYFDKLVHGAANILQCLFGNNADTTVAILYNFLNVPYKLFS"
     gene       15429. .16679
                /gene="ORF10"
     CDS        15429. .16679
                /gene="ORF10"
                /codon_start=1
                /product="ORF10"
                /protein_id="AAF59987.1"
                /db_xref="GI:7329998"
                /translation="MLVNELSVVLGDWEVTFHRGRFSFVNLTRLQTFKGHGGYARVRL
                PFSLDQLLHQHFAFGLVTRLKELPPFSDCVALIAPLDSGGDADAARVAPGFVLDSSRP
                LTVWVNASGRHTIRFCLLFLKPIDLERAVTYVFGENGGARSEGTPKPTCATESLPGGP
                LRVSGEASQTSPHSFVAYFPTANSVACLSLLRLQVRPFSDDAAHRDARISPKYVTFSN
                SGGNVCKASVHTLSPSRCKTAQMEIIYAPGDPNAEIVLGQSGPVLPTHTGGRVLGVYA
                DAEKTIQPGSSAEVRVQLIFQQGAAARGDLAFLVTGVAPEPLFVVTPALLLSGCTTHL
                RLFNPNGTPTTIKRDTLVAAAAPCPVVRLSSADDAPRDLVASPDTGALSINAFTIPVG
                FPGVVSAECHVSLRDNGVHERMNH"
     gene       16688. .17917
                /gene="ORF11"
     CDS        16688. .17917
                /gene="ORF11"
                /codon_start=1
                /product="ORF11"
                /protein_id="AAF59988.1"
                /db_xref="GI:7329999"
                /translation="MGTPVRFFRGEWQTSSLVDNGTPRYSSLVWAATIHDGYLTLVNR
                SELCVTERSPCLPACPSIGRLVGKRFPGFAFASATLGDRGTRTVFYAFGHRDNPLDIV
                PAVVERADRELVLRVHAPQTTRVSRYGLKVFVAIVTVVRPPGVFLHFPQDRVPIALTD

  Query Match             23.9%;  Score 239.2;  DB 14;  Length 130733;
  Best Local Similarity   88.1%;  Pred. No. 2.9e-29;
  Matches  273;  Conservative    0;  Mismatches   33;  Indels    4;  Gaps    1;

Qy        219 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 278
              | |||| ||| |||||||||||||||||||||     |   |   || ||||||||||||
Db     130428 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 130487

Qy        279 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 338
              ||||||||||||     |   |  ||||||||||||     |   |   || |||||||||
Db     130488 GCCTGCTTGCT----TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 130543

Qy        339 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 398
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     130544 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 130603

Qy        399 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 458
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     130604 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 130663

Qy        459 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 518
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     130664 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 130723

Qy        519 GCCTGCTTGC 528
              ||||||||||
Db     130724 GCCTGCTTGC 130733


RESULT 6
AC097912/c
LOCUS       AC097912               218921 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-154L5, *** SEQUENCING IN PROGRESS
            ***.
ACCESSION   AC097912
VERSION     AC097912.5  GI:30520994
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 218921)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 218921)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 218921)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:22855232.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
```

```
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GFOX
                Center clone name: CH230-154L5
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 202489 bases at least Q40
                Consensus quality: 204424 bases at least Q30
                Consensus quality: 205792 bases at least Q20
                Estimated insert size: 214534; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   218921: contig of 218921 bp in length.
FEATURES             Location/Qualifiers
     source          1. .218921
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-154L5"
     misc_feature    1. .1273
                     /note="wgs_contig"
BASE COUNT    57277 a  42734 c  45435 g  61021 t  12454 others
ORIGIN

  Query Match             22.8%;  Score 228;  DB 2;  Length 218921;
  Best Local Similarity   62.2%;  Pred. No. 1.5e-27;
  Matches  377;  Conservative    0;  Mismatches  225;  Indels    4;  Gaps    1;

Qy        240 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 299
              | ||  | | |||    ||||  |||||||| |  ||| ||||||||||      |   |
Db      89107 CATGTCTTCAGAGTTTTCAGTTTGCCTGCTTACCTGCTTGCCTGCTTGCCTGCCTGCTTG 89048

Qy        300 TAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 359
                || ||| |||      | | |   ||||||||||  ||| || ||| ||||     |
Db      89047 CCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCTTGCCTGCTTGCTTG 88988

Qy        360 ACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 419
              | |   |  |||| || |||| |   |  |||  ||| ||||||||||| |||| ||  |
Db      88987 CCTGCTTGCTTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTG 88928

Qy        420 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 479
              | ||| |||||||  ||| ||||||  |||||||  ||| ||||||  |||||||  ||| |
Db      88927 CTTGCCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTGCTTG 88868

Qy        480 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 539
              |||||||||  |||| ||  ||||||||||  |||| ||  ||||||||||||||  ||| |
Db      88867 CCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTGCCTGCTTGCTTGCCTGCTTG 88808

Qy        540 CCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTG 599
              |||||||||      | | |   |  ||| ||| ||||     | | |   |||||||||
Db      88807 CCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTGCCTGCTTGCCTGCCTGCTTGCCTGCCTG 88748

Qy        600 ATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 659
                  ||||        | ||||  |  |||  |   | |||||  |  |||  |   | |||
Db      88747 CCTACTTGCCTACTTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCTTGCCTGCTTGCCTG 88688

Qy        660 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTG 719
              || | |||||  |   |  |||| | |||||  |     || || | ||||  |   | ||
Db      88687 CTTGCCTGCTTGCCTGCTTGCTTGCCTGCTTGC----CTGCTTGCCTGCCTGCTTGCCTA 88632

Qy        720 CTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 779
              || | ||||  |   | || || | |||||  |   | |||||  | |||||  |   | ||
Db      88631 CTTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTGCTTG 88572

Qy        780 CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTG 839
              || | ||||  |   |  ||| |  ||||     |  | | || |  |||||||  ||||
Db      88571 CTTGCCTGCCTGCCTGCTTGCCTGCTTGCTTCCTTGTTTCCTTGATATCTTGCTTCCTTG 88512

Qy        840 CTTGCT 845
              |||| |
Db      88511 CTTGGT 88506


RESULT 7
AC130110/c
LOCUS       AC130110               229535 bp    DNA     linear   HTG 22-SEP-2002
DEFINITION  Rattus norvegicus clone CH230-55M4, *** SEQUENCING IN PROGRESS ***,
            2 unordered pieces.
ACCESSION   AC130110
VERSION     AC130110.2  GI:23121680
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 229535)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
```

```
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 229535)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 229535)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 18, 2002 this sequence version replaced gi:22138333.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequening reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/).  As a result, the
            sequence may extend beyond the ends of the clone and there may be
            contigs that consist entirely of whole genome shotgun sequence
            reads.  Both end sequences and whole genome shotgun sequence only
            contigs will be indicated in the feature table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GJJP
                Center clone name: CH230-55M4
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 189188 bases at least Q40
                Consensus quality: 193629 bases at least Q30
                Consensus quality: 196703 bases at least Q20
                Estimated insert size: 207604; sum-of-contigs estimation
                Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   192206: contig of 192206 bp in length
            *  192207   192306: gap of unknown length
            *  192307   229535: contig of 37229 bp in length.
FEATURES            Location/Qualifiers
     source         1. .229535
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-55M4"
     misc_feature   48048. .49205
                    /note="wgs_contig"
     misc_feature   49707. .51011
                    /note="wgs_contig"
     misc_feature   complement(189833. .190293)
                    /note="clone_boundary
                    clone_end:T7
                    site:EcoRI
                    end_sequence:BH266273"
BASE COUNT    52992 a  49371 c  47396 g  49177 t  30599 others
ORIGIN

  Query Match             22.4%;  Score 224;  DB 2;  Length 229535;
  Best Local Similarity   58.3%;  Pred. No. 6.5e-27;
  Matches  392;  Conservative    0;  Mismatches  280;  Indels    0;  Gaps    0;

Qy        194 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 253
              ||| || |||   |    |  |   || ||  ||| || | ||| |  || ||| |  |
Db     211432 TGCTTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGCTTGGTTGCTTGGTTGTTTGGTTGGT 211373

Qy        254 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 313
              |   |   | |||||| | ||| | ||||||||  |  |   |   |  || ||| |
Db     211372 TGCTTGGTTGGTTGCTTGGTTGCTTGGTTGCTTGCTTGGTTGCTTGCTTGGTTGGTTGTT 211313

Qy        314 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 373
              |  |   |   | || ||||| | | |  || ||| |      |   |   |  || |
Db     211312 TGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGCTTGGT 211253

Qy        374 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 433
              | |||||  |   || |||| | | |  |||||| ||||| | | || || ||||||||||
Db     211252 TGGTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTTGCTTGGTTGCTTGCT 211193

Qy        434 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 493
              | | |  || |||||||  | | | || |||||| ||||| |||  |  |||||| |||||
Db     211192 TGGTTGGTTGGTTGCTTGGTTGGTTGCTTGCTTGGTTGCTTGCTTGGTTGCTTGGTTGCT 211133

Qy        494 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGG 553
              |||  |   || ||| ||| | | | |  |||||  |||  | ||| || ||||||||
Db     211132 TGCTTGGTTGGTTGGTTGGTTGGTTGCTTAGTTGGTTGCTTGCTTGCTTGCTTGCT 211073

Qy        554 GGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTA 613
              |   |   |  |||||| |||||  |  |   |   |  ||| || | |||| || |
Db     211072 TGGTTGGTTGCTTGCTTGGTTGCTTGGTTGTTTGGTTGGTTGCTTGGTTGCTTCATTGCT 211013

Qy        614 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 673
              || || | || || | |   || || | || |||| |   |  |||| |  |||| |
Db     211012 TGGTTGGTTGGTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 210953

Qy        674 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 733
              | || | |  || | |   || |  | || || | |   || || | |  |||| |
Db     210952 TAGTTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGGTTGCTTGCTTGCT 210893

Qy        734 GGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 793
              || | || || || | |   || |||| || |||| |   || |||| || |||| |
Db     210892 TGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTTGCTTGCTTGGTTGCTTGGTTGCTTGCT 210833

Qy        794 GGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGC 853
              || |||| |  |||| |   | ||| |||| | |||||||||| ||||| |   |
Db     210832 TGGTTGCTTGCTTGCTTGCTTGCTTGCTTGGTTGCTTGCTTGGTTGCTTGGTTGCT 210773

Qy        854 CGCTTGCCTGCT 865
              | |||| || |
Db     210772 TGGTTGCTTGGT 210761


RESULT 8
AC137146/c
LOCUS       AC137146               208050 bp    DNA     linear   HTG 08-MAR-2003
DEFINITION  Mus musculus clone RP24-358I14, WORKING DRAFT SEQUENCE, 16
            unordered pieces.
ACCESSION   AC137146
VERSION     AC137146.2  GI:28882257
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 208050)
```

```
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-358I14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 208050)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 208050)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 8, 2003 this sequence version replaced gi:25048829.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L25975
                Center clone name: 358_I_14
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 203377 bases at least Q40
                Consensus quality: 205277 bases at least Q30
                Consensus quality: 206093 bases at least Q20
                Insert size: 188000; agarose-fp
                Insert size: 206550; sum-of-contigs
                Quality coverage: 10.1 in Q20 bases; agarose-fp
                Quality coverage: 9.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1438: contig of 1438 bp in length
            *     1439     1538: gap of 100 bp
            *     1539     2410: contig of 872 bp in length
            *     2411     2510: gap of 100 bp
            *     2511     3868: contig of 1358 bp in length
            *     3869     3968: gap of 100 bp
            *     3969     5765: contig of 1797 bp in length
            *     5766     5865: gap of 100 bp
            *     5866     7076: contig of 1211 bp in length
            *     7077     7176: gap of 100 bp
            *     7177     8356: contig of 1180 bp in length
            *     8357     8456: gap of 100 bp
            *     8457     9860: contig of 1404 bp in length
            *     9861     9960: gap of 100 bp
            *     9961    11098: contig of 1138 bp in length
            *    11099    11198: gap of 100 bp
            *    11199    13131: contig of 1933 bp in length
            *    13132    13231: gap of 100 bp
            *    13232    15195: contig of 1964 bp in length
            *    15196    15295: gap of 100 bp
            *    15296    19695: contig of 4400 bp in length
            *    19696    19795: gap of 100 bp
            *    19796    24098: contig of 4303 bp in length
            *    24099    24198: gap of 100 bp
            *    24199    31817: contig of 7619 bp in length
            *    31818    31917: gap of 100 bp
            *    31918    44002: contig of 12085 bp in length
            *    44003    44102: gap of 100 bp
            *    44103   110199: contig of 66097 bp in length
            *   110200   110299: gap of 100 bp
            *   110300   208050: contig of 97751 bp in length.
FEATURES             Location/Qualifiers
     source          1. .208050
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-358I14"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .1438
                     /note="assembly_fragment"
     misc_feature    1539. .2410
                     /note="assembly_fragment"
     misc_feature    2511. .3868
                     /note="assembly_fragment"
     misc_feature    3969. .5765
                     /note="assembly_fragment"
     misc_feature    5866. .7076
                     /note="assembly_fragment"
     misc_feature    7177. .8356
                     /note="assembly_fragment"
     misc_feature    8457. .9860
                     /note="assembly_fragment"
     misc_feature    9961. .11098
                     /note="assembly_fragment"
     misc_feature    11199. .13131
                     /note="assembly_fragment"
     misc_feature    13232. .15195
                     /note="assembly_fragment"
     misc_feature    15296. .19695
                     /note="assembly_fragment"
     misc_feature    19796. .24098
                     /note="assembly_fragment"
     misc_feature    24199. .31817
                     /note="assembly_fragment"
     misc_feature    31918. .44002
```

```
                     /note="assembly_fragment"
     misc_feature    44103. .110199
                     /note="assembly_fragment"
     misc_feature    110300. .208050
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    58437 a  49871 c  45960 g  52282 t   1500 others
ORIGIN

  Query Match           21.9%;  Score 219;  DB 2;  Length 208050;
  Best Local Similarity   57.5%;  Pred. No. 4.1e-26;
  Matches  393;  Conservative    0;  Mismatches  290;  Indels    0;  Gaps    0;

Qy        194 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 253
              ||| ||| ||   | |  |  |   | ||||||||||||  ||  |||||  |||
Db     161257 TGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCC 161198

Qy        254 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 313
              | | |   ||||||| |||  ||  |||||| |||       | | |   |||||| ||
Db     161197 TGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTC 161138

Qy        314 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 373
                  | |  ||  |||||| |||  |   |||||| |||     | | ||  | |||
Db     161137 TGCCTGCCTGTCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGTC 161078

Qy        374 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 433
              ||  |||  |   | |||   ||  |||||| ||| ||   ||  |||||| ||| |||
Db     161077 TGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCC 161018

Qy        434 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 493
              ||  ||||||  ||| |||  ||  |||||| ||| |||  |   |||||| ||| |||
Db     161017 TGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCC 160958

Qy        494 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGG 553
              |   |||||| ||| |||  |   |||||| ||| |||  |   |||||| ||
Db     160957 TGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTGCC 160898

Qy        554 GGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTA 613
              | | |   | |||| ||| |||      | | |   | |||||||   || ||   ||
Db     160897 TGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGTC 160838

Qy        614 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 673
              | ||||  | ||||  ||  | ||||  | ||||  ||  | ||||  | ||||  |
Db     160837 TGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTGCCTGCC 160778

Qy        674 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 733
              | ||||  | ||||  ||  | ||  |  | ||||  |   | ||||  | |||  |
Db     160777 TGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTGCC 160718

Qy        734 GGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 793
              | ||  |  | |||   |   | ||||  | |||   |   | ||||  | ||||  ||
Db     160717 TGCCTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTC 160658

Qy        794 GGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGC 853
              | ||||  | ||||  |   | ||||||||  || ||| ||| ||| |||  |   | |
Db     160657 TGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTC 160598

Qy        854 CGCTTGCCTGCTACTAGGGCTGC 876
              || |||||||      | ||||
Db     160597 TGCCTGCCTGCCTGCCTGCTGC 160575


RESULT 9
AC112675/c
LOCUS       AC112675              228919 bp    DNA     linear   HTG 21-OCT-2002
DEFINITION  Mus musculus clone RP23-138J12, WORKING DRAFT SEQUENCE, 7 ordered
            pieces.
ACCESSION   AC112675
VERSION     AC112675.4  GI:24182124
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 228919)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-138J12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 228919)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 228919)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 21, 2002 this sequence version replaced gi:21070725.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L22426
                Center clone name: 138_J_12
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 227366 bases at least Q40
                Consensus quality: 227820 bases at least Q30
                Consensus quality: 228087 bases at least Q20
```

```
            Insert size: 230000; agarose-fp
            Insert size: 228319; sum-of-contigs
            Quality coverage: 12.8 in Q20 bases; agarose-fp
            Quality coverage: 12.9 in Q20 bases; sum-of-contigs
          ----------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 7 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *        1    48936: contig of 48936 bp in length
          *    48937    49036: gap of 100 bp
          *    49037    50559: contig of 1523 bp in length
          *    50560    50659: gap of 100 bp
          *    50660    56278: contig of 5619 bp in length
          *    56279    56378: gap of 100 bp
          *    56379    90153: contig of 33775 bp in length
          *    90154    90253: gap of 100 bp
          *    90254   114936: contig of 24683 bp in length
          *   114937   115036: gap of 100 bp
          *   115037   207735: contig of 92699 bp in length
          *   207736   207835: gap of 100 bp
          *   207836   228919: contig of 21084 bp in length.
FEATURES             Location/Qualifiers
     source          1. .228919
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-138J12"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .48936
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    49037. .50559
                     /note="assembly_fragment"
     misc_feature    50660. .56278
                     /note="assembly_fragment"
     misc_feature    56379. .90153
                     /note="assembly_fragment"
     misc_feature    90254. .114936
                     /note="assembly_fragment"
     misc_feature    115037. .207735
                     /note="assembly_fragment"
     misc_feature    207836. .228919
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    64052 a  55854 c  51189 g  57224 t    600 others
ORIGIN

  Query Match             21.9%;  Score 219;  DB 2;  Length 228919;
  Best Local Similarity   57.5%;  Pred. No. 4e-26;
  Matches  393;  Conservative    0;  Mismatches  290;  Indels    0;  Gaps    0;

Qy        194 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 253
              ||| ||| ||   | |  |  |   | ||||||||||||  ||  |||||  |||
Db        782 TGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCC 723

Qy        254 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 313
              | | |   ||||||  |||  ||  ||||||  |||      | | |   ||||||  ||
Db        722 TGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTC 663

Qy        314 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 373
                  | | ||  ||||||  |||  |   ||||||  |||       | | ||  | |||
Db        662 TGCCTGCCTGTCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGTC 603

Qy        374 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 433
              | || |||  |   | |||   ||  ||||||  |||  ||   ||  ||||||  |||  |||

Db        602 TGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCC 543

Qy        434 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 493
              ||  ||||||  |||  |||  ||  ||||||  |||  |||  |   ||||||  |||  |||
Db        542 TGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCC 483

Qy        494 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGG 553
              |   ||||||  |||  |||  |   ||||||  |||  |||  |   ||||||  ||
Db        482 TGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTGCC 423

Qy        554 GGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTA 613
              | | |   | ||||  |||  |||       | | |   | |||||||   ||  ||   ||
Db        422 TGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGTC 363

Qy        614 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 673
              | ||||  | ||||  ||  | ||||  | ||||  ||  | ||||  | ||||  |
Db        362 TGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTGCCTGCC 303

Qy        674 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 733
              | ||||  | ||||  ||  | ||  | ||||  |   | ||||  | |||  |
Db        302 TGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTGCC 243

Qy        734 GGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 793
              | ||  |  |||   |   | ||||  | |||   |   | ||||  | ||||  ||
Db        242 TGCCTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTC 183

Qy        794 GGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGC 853
              | ||||  | ||||  |   | ||||||||  ||  |||  |||  |||  |||  |   | |
Db        182 TGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTC 123

Qy        854 CGCTTGCCTGCTACTAGGGCTGC 876
              ||  |||||||       | ||||
Db        122 TGCCTGCCTGCCTGCCTGCCTGC 100


RESULT 10
AC131862/c
LOCUS       AC131862              245853 bp    DNA     linear   HTG 10-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-10F20, *** SEQUENCING IN PROGRESS
            ***.
ACCESSION   AC131862
VERSION     AC131862.3  GI:23674277
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 245853)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
```

```
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 245853)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 245853)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 10, 2002 this sequence version replaced gi:22539029.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GDSM
                Center clone name: CH230-10F20
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 227608 bases at least Q40
                Consensus quality: 230076 bases at least Q30
                Consensus quality: 231849 bases at least Q20
                Estimated insert size: 235882; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   245853: contig of 245853 bp in length.
FEATURES            Location/Qualifiers
     source         1. .245853
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-10F20"
     misc_feature   1. .1017
                    /note="wgs_contig"
     misc_feature   1068. .2461
                    /note="wgs_contig"
     misc_feature   2512. .3853
                    /note="wgs_contig"
     misc_feature   6786. .8519
                    /note="wgs_contig"
BASE COUNT    68006 a  49515 c  49070 g  66927 t  12335 others
ORIGIN

  Query Match             21.6%;  Score 216.6;  DB 2;  Length 245853;
  Best Local Similarity   63.8%;  Pred. No. 9.5e-26;
  Matches  345;  Conservative    0;  Mismatches  194;  Indels    2;  Gaps    1;

Qy        327 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 386
              ||||||||||  ||| || ||||||||     |   |   | ||||| ||||||  |
Db     175578 GCCTGCTTGCCTGCTTGCTTGCTTGCTTGCCTGCTTGCTTGCCTGCTTGCTTGCCTGCTT 175519

Qy        387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG--CTTGCTTGCTCGCTGGCCTGC 444
              | ||||| ||| || ||| ||| |||| ||| || ||  ||||| |||  ||  ||||||
Db     175518 GCCTGCTTGCTTGCTTGCCTGCCTGCTTGCTTGCTTGCTCTTGCCTGCCTGCCTGCCTGC 175459

Qy        445 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 504
              |||||||  ||  || ||||||||||| ||| ||||||||| |||  ||  ||||||
Db     175458 CTGCTTGCCTGCCTGCTTGCTTGCTTGCTTGCCTGCTTGCCTGCCTGCCTGCCTGC 175399

Qy        505 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGG 564
              ||||||||| ||| || ||||||||||| ||| ||||||  |||     |   |   |
Db     175398 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGCCTGCCTGCCTGCTTGCTTGC 175339

Qy        565 CTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGG 624
              ||||||||||||||     |   |   |||| |||| | ||||| | |   | ||||  |
Db     175338 CTGCTTGCTTGCTTGCTTGCTTGCTTGCCTACCTGCTTGCTTGCTTGCTTGCCTGCCTGC 175279

Qy        625 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 684
              ||||  |   | ||||| |  |||| |   |  | | |  |  |||| |   |  |||| |
Db     175278 CTGCCTGCCTGCCTGCTTGCTTGCTTGCTTACCTGCTTGCTTGCCTGCTTGCTTGC 175219

Qy        685 CTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGG 744
              ||||  |   |  | |  | ||||  |   | ||||  | ||||  |   | || |  |
Db     175218 TTGCTTGCTTGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGC 175159

Qy        745 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 804
              ||||  |   | |||||  | |||||  |   |  |||| | |||||  |   |  |||| |
Db     175158 CTGCCTGCCTGCCTGCTTGCCTGCTTGCTCGCTTGCTCGCCTGCTTGCCTGCTTGCTTGC 175099

Qy        805 CTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGC 864
              ||||  |   |  ||| |||| |||||| ||| ||| ||| |   |  |||| ||  ||
Db     175098 TTGCTTGCTTGCTTGCTTGCTTGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGC 175039

Qy        865 T 865
              |
Db     175038 T 175038


RESULT 11
AC106141
LOCUS       AC106141              305166 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-229F12, *** SEQUENCING IN PROGRESS
```

```
            ***, 17 unordered pieces.
ACCESSION   AC106141
VERSION     AC106141.4  GI:30579567
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 305166)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 305166)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 305166)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:23664597.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
               Center project name: GJGS
               Center clone name: CH230-229F12
            -------------- Summary Statistics
               Assembly program: Atlas 3.0;
               Consensus quality: 278355 bases at least Q40
               Consensus quality: 283506 bases at least Q30
               Consensus quality: 287188 bases at least Q20
               Estimated insert size: 285546; sum-of-contigs estimation
               Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3937: contig of 3937 bp in length
            *     3938     4037: gap of unknown length
            *     4038     8398: contig of 4361 bp in length
            *     8399     8498: gap of unknown length
            *     8499    11821: contig of 3323 bp in length
            *    11822    11921: gap of unknown length
            *    11922    27313: contig of 15392 bp in length
            *    27314    27413: gap of unknown length
            *    27414    31822: contig of 4409 bp in length
            *    31823    31922: gap of unknown length
            *    31923   272126: contig of 240204 bp in length
            *   272127   272226: gap of unknown length
            *   272227   275425: contig of 3199 bp in length
            *   275426   275525: gap of unknown length
            *   275526   282534: contig of 7009 bp in length
            *   282535   282634: gap of unknown length
            *   282635   283878: contig of 1244 bp in length
            *   283879   283978: gap of unknown length
            *   283979   285225: contig of 1247 bp in length
            *   285226   285325: gap of unknown length
            *   285326   286582: contig of 1257 bp in length
            *   286583   286682: gap of unknown length
            *   286683   288078: contig of 1396 bp in length
            *   288079   288178: gap of unknown length
            *   288179   289887: contig of 1709 bp in length
            *   289888   289987: gap of unknown length
            *   289988   291764: contig of 1777 bp in length
            *   291765   291864: gap of unknown length
            *   291865   295767: contig of 3903 bp in length
            *   295768   295867: gap of unknown length
            *   295868   299271: contig of 3404 bp in length
            *   299272   299371: gap of unknown length
            *   299372   305166: contig of 5795 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1. .305166
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-229F12"
     misc_feature    1. .1008
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    2758. .3559
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:RWBOD30TVB"
     misc_feature    23557. .25155
                     /note="wgs_contig"
     misc_feature    31923. .33139
                     /note="wgs_contig"
     misc_feature    237160. .237846
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:RWBOD30TJB"
     misc_feature    271133. .272126
                     /note="wgs_end_extension
                     clone_end:T7"
BASE COUNT    88552 a  60606 c  60203 g  80778 t  15027 others
ORIGIN

  Query Match             21.6%;  Score 216.6;  DB 2;  Length 305166;
  Best Local Similarity   63.8%;  Pred. No. 8.9e-26;
  Matches  345;  Conservative    0;  Mismatches  194;  Indels    2;  Gaps    1;

Qy        327 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 386
              ||||||||||  ||| || ||||||||     |   |   | |||||  ||||||  |
Db     247507 GCCTGCTTGCCTGCTTGCTTGCTTGCTTGCCTGCTTGCTTGCCTGCTTGCTTGCCTGCTT 247566

Qy        387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG--CTTGCTTGCTCGCTGGCCTGC 444
              | |||||  ||| ||  |||  |||  ||||  ||| || ||  |||||  |||  ||  ||||||
Db     247567 GCCTGCTTGCTTGCTTGCCTGCCTGCTTGCTTGCTTGCTCTTGCCTGCCTGCCTGCCTGC 247626

Qy        445 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 504
              |||||||  ||  || ||||||||||||  ||| ||||||||||  |||  ||  ||||||
Db     247627 CTGCTTGCCTGCCTGCTTGCTTGCTTGCTTGCCTGCTTGCCTGCCTGCCTGCCTGCTGC 247686

Qy        505 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGG 564
              |||||||||  ||| ||  ||||||||||||  ||| ||||||  |||      |   |   |
Db     247687 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGCCTGCCTGCCTGCTTGCTTGC 247746

Qy        565 CTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGG 624
              ||||||||||||||      |   |   |||| |||| | |||||  | |   | ||||  |
Db     247747 CTGCTTGCTTGCTTGCTTGCTTGCTTGCCTACCTGCTTGCTTGCTTGCTTGCCTGCCTGC 247806

Qy        625 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 684
              ||||  |   | |||||  |  ||||  |   | |  |  ||||  |   |  |||| |
Db     247807 CTGCCTGCCTGCCTGCTTGCTTGCTTGCTTGCTTACCTGCTTGCTTGCCTGCTTGCTTGC 247866

Qy        685 CTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGG 744
              ||||  |   |  | |  | ||||  |   | ||||  | ||||  |   |  || |  |
Db     247867 TTGCTTGCTTGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGC 247926

Qy        745 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 804
              ||||  |   | |||||  | |||||  |   |  |||| | |||||  |   |  |||| |
Db     247927 CTGCCTGCCTGCCTGCTTGCCTGCTTGCTCGCTTGCTCGCCTGCTTGCCTGCTTGCTTGC 247986

Qy        805 CTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGC 864
              ||||  |   |  ||| |||| |||||||  ||| ||| ||| |   |  |||| ||  ||
Db     247987 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGC 248046

Qy        865 T 865
              |
Db     248047 T 248047
```

```
RESULT 12
AC115409
LOCUS       AC115409               220464 bp    DNA     linear   HTG 13-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-134D4, WORKING DRAFT SEQUENCE.
ACCESSION   AC115409
VERSION     AC115409.4  GI:24941929
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 220464)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 220464)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 220464)
  AUTHORS   Rat Genome Sequencing Consortium.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 13, 2002 this sequence version replaced gi:22857077.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GOMT
                Center clone name: CH230-134D4
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 207904 bases at least Q40
                Consensus quality: 209884 bases at least Q30
                Consensus quality: 210779 bases at least Q20
                Estimated insert size: 213748; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   220464: contig of 220464 bp in length.
FEATURES             Location/Qualifiers
     source          1. .220464
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-134D4"
     misc_feature    1. .2329
                     /note="wgs_contig"
     misc_feature    219334. .220464
                     /note="wgs_contig"
BASE COUNT    60311 a  46976 c  47584 g  56550 t   9043 others
ORIGIN

  Query Match             21.0%;  Score 210.4;  DB 2;  Length 220464;
  Best Local Similarity   57.2%;  Pred. No. 9.4e-25;
  Matches  382;  Conservative    0;  Mismatches  286;  Indels    0;  Gaps    0;

Qy        209 CACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGC 268
              ||||  |   | ||||||||||||  ||  |||||| |||         | | . ||||||
Db      81443 CACCATGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTCCCTGCCTGCCTGC 81502

Qy        269 TTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGC 328
              ||   ||  |||||| |||      | | ||  |||||| |||      | | ||  |
Db      81503 CTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGT 81562

Qy        329 CTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGG 388
              |||| |||  |   |||||  |||      | | |   | ||||  || |||  |   |
Db      81563 CTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCTGC 81622

Qy        389 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 448
              ||||  |  ||||||| ||| |||  |   ||||||| ||| |||  ||  | |||| |||
Db      81623 CTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGC 81682

Qy        449 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 508
              |||  ||  | |||| ||| ||   ||  |||||| ||| ||   ||  |||||| |||
Db      81683 CTGCCTGCCTGTCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGC 81742

Qy        509 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGC 568
              ||   ||  |||||| ||  |||  ||  |||||| ||        | | |   | ||||
Db      81743 CTGTCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGC 81802

Qy        569 TTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGC 628
              |||  |||       | | ||  | |||||||   |  ||   ||  | ||||  | ||||
Db      81803 CTGCCTGCCTGCCTGCCTGTCTGTCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGTCTGC 81862

Qy        629 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 688
              |   | ||||  | |||   |   | ||||  | ||||  ||  | ||||  | ||||
Db      81863 CTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGC 81922

Qy        689 TAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGC 748
              ||  | ||  |  | ||||  ||  | ||||  | |||   |   | ||    | ||||
Db      81923 CTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTGCCTGC 81982

Qy        749 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 808
              |   | ||||  | |||   ||  | ||||  | ||||  ||  | ||||  | ||||
Db      81983 CTGCCTGCCTGCCTGCCTGTCTGTCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGTCTGC 82042

Qy        809 TAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACT 868
              |   | ||||||||  |  ||| ||| ||| |||  |   | | || |||||||   |
Db      82043 CTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGT 82102

Qy        869 AGGGCTGC 876
              | || |
Db      82103 CTGCCTAC 82110


RESULT 13
AC141883
LOCUS       AC141883               238215 bp    DNA     linear   HTG 20-MAR-2003
DEFINITION  Mus musculus chromosome UNK clone RP24-128P19, WORKING DRAFT
            SEQUENCE, 31 unordered pieces.
ACCESSION   AC141883
VERSION     AC141883.2  GI:29126550
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 238215)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 238215)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 238215)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Mar 20, 2003 this sequence version replaced gi:29124284.

            ---------------- Genome Center ----------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ------------- Project Information ------------------
            Center project name: M_BB0128P19
```

```
----------------------------------------------------
-------------- Summary Statistics  --------------
Sequencing vector: M13; %
Sequencing vector: plasmid; %
Chemistry: Dye-primer ET; % of reads
Chemistry: Dye-terminator Big Dye; % of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 221212 bases at least Q40
Consensus quality: 224001 bases at least Q30
Consensus quality: 225617 bases at least Q20
----------------------------------------------------
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1212: contig of 1212 bp in length
*     1213     1312: gap of unknown length
*     1313     2808: contig of 1496 bp in length
*     2809     2908: gap of unknown length
*     2909     4509: contig of 1601 bp in length
*     4510     4609: gap of unknown length
*     4610     6126: contig of 1517 bp in length
*     6127     6226: gap of unknown length
*     6227     7673: contig of 1447 bp in length
*     7674     7773: gap of unknown length
*     7774     8968: contig of 1195 bp in length
*     8969     9068: gap of unknown length
*     9069    10583: contig of 1515 bp in length
*    10584    10683: gap of unknown length
*    10684    11778: contig of 1095 bp in length
*    11779    11878: gap of unknown length
*    11879    13464: contig of 1586 bp in length
*    13465    13564: gap of unknown length
*    13565    14971: contig of 1407 bp in length
*    14972    15071: gap of unknown length
*    15072    16342: contig of 1271 bp in length
*    16343    16442: gap of unknown length
*    16443    17795: contig of 1353 bp in length
*    17796    17895: gap of unknown length
*    17896    19947: contig of 2052 bp in length
*    19948    20047: gap of unknown length
*    20048    21207: contig of 1160 bp in length
*    21208    21307: gap of unknown length
*    21308    23316: contig of 2009 bp in length
*    23317    23416: gap of unknown length
*    23417    24878: contig of 1462 bp in length
*    24879    24978: gap of unknown length
*    24979    26136: contig of 1158 bp in length
*    26137    26236: gap of unknown length
*    26237    27558: contig of 1322 bp in length
*    27559    27658: gap of unknown length
*    27659    29825: contig of 2167 bp in length
*    29826    29925: gap of unknown length
*    29926    31576: contig of 1651 bp in length
*    31577    31676: gap of unknown length
*    31677    33353: contig of 1677 bp in length
*    33354    33453: gap of unknown length
*    33454    35537: contig of 2084 bp in length
*    35538    35637: gap of unknown length
*    35638    37238: contig of 1601 bp in length
*    37239    37338: gap of unknown length
*    37339    38481: contig of 1143 bp in length
*    38482    38581: gap of unknown length
*    38582    40522: contig of 1941 bp in length
*    40523    40622: gap of unknown length
*    40623    45987: contig of 5365 bp in length
*    45988    46087: gap of unknown length
*    46088    70413: contig of 24326 bp in length
*    70414    70513: gap of unknown length
*    70514   101243: contig of 30730 bp in length
*   101244   101343: gap of unknown length
*   101344   136512: contig of 35169 bp in length
*   136513   136612: gap of unknown length
*   136613   183419: contig of 46807 bp in length
*   183420   183519: gap of unknown length
*   183520   238215: contig of 54696 bp in length.
FEATURES             Location/Qualifiers
     source          1. .238215
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP24-128P19"
     misc_feature    1. .1212
                     /note="assembly_name:Contig13"
     misc_feature    1313. .2808
                     /note="assembly_name:Contig14"
     misc_feature    2909. .4509
                     /note="assembly_name:Contig16"
     misc_feature    4610. .6126
                     /note="assembly_name:Contig17"
     misc_feature    6227. .7673
                     /note="assembly_name:Contig18"
     misc_feature    7774. .8968
                     /note="assembly_name:Contig19"
     misc_feature    9069. .10583
                     /note="assembly_name:Contig20"
     misc_feature    10684. .11778
                     /note="assembly_name:Contig26"
     misc_feature    11879. .13464
                     /note="assembly_name:Contig27"
     misc_feature    13565. .14971
                     /note="assembly_name:Contig28"
     misc_feature    15072. .16342
                     /note="assembly_name:Contig30"
     misc_feature    16443. .17795
                     /note="assembly_name:Contig31"
     misc_feature    17896. .19947
                     /note="assembly_name:Contig33"
     misc_feature    20048. .21207
                     /note="assembly_name:Contig34"
     misc_feature    21308. .23316
                     /note="assembly_name:Contig35"
     misc_feature    23417. .24878
                     /note="assembly_name:Contig37"
     misc_feature    24979. .26136
                     /note="assembly_name:Contig38"
     misc_feature    26237. .27558
                     /note="assembly_name:Contig40"
     misc_feature    27659. .29825
                     /note="assembly_name:Contig41"
     misc_feature    29926. .31576
                     /note="assembly_name:Contig42"
     misc_feature    31677. .33353
                     /note="assembly_name:Contig43"
     misc_feature    33454. .35537
                     /note="assembly_name:Contig44"
     misc_feature    35638. .37238
                     /note="assembly_name:Contig45"
     misc_feature    37339. .38481
                     /note="assembly_name:Contig46"
     misc_feature    38582. .40522
                     /note="assembly_name:Contig47"
     misc_feature    40623. .45987
                     /note="assembly_name:Contig48"
     misc_feature    46088. .70413
                     /note="assembly_name:Contig49"
     misc_feature    70514. .101243
                     /note="assembly_name:Contig50"
     misc_feature    101344. .136512
                     /note="assembly_name:Contig51"
     misc_feature    136613. .183419
```

```
                     /note="assembly_name:Contig52"
     misc_feature    183520. .238215
                     /note="assembly_name:Contig53"
BASE COUNT    62298 a  54994 c  55470 g  62353 t   3100 others
ORIGIN

  Query Match             21.0%;  Score 209.8;  DB 2;  Length 238215;
  Best Local Similarity   62.9%;  Pred. No. 1.1e-24;
  Matches  325;  Conservative    0;  Mismatches  192;  Indels    0;  Gaps    0;

Qy        329 CTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGG 388
              ||| ||| | ||| |||||||||| |     | | |   | |||  ||||| | |   |
Db     102066 CTGTTTGTTTGCTTGCCTGCTTGTTTGCTTGCCTGCCTGCTTGCCTGCTTGTTTGCTTGC 102125

Qy        389 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 448
              ||||| | | || ||  ||| |||  ||| |||||| || ||||| ||  || || ||||
Db     102126 CTGCTTGTTTGCTTGTCTGCCTGCCTGCTTGCCTGCCTGTTTGCTTGCCTGCTTGTTTGC 102185

Qy        449 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 508
              ||||  ||| |  ||||||| |||| ||| ||||||||||||||  |   |||||| |||
Db     102186 TTGCCTGCTTGTTTGCTTGCCTGCTTGCTTGCCTGCTTGCTTGCCTGACTGCCTGCCTGC 102245

Qy        509 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGC 568
              ||||| ||| || ||| |||||||  ||| ||||||||||       | | |   | ||||
Db     102246 TTGCTTGCTTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGC 102305

Qy        569 TTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGC 628
              |||| ||||     |   |  |||||||||  || ||   |   | |||||  |  |||
Db     102306 TTGCCTGCTTGCCTGCTTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGTTTGC 102365

Qy        629 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 688
              | |   |  |||  | |||  |   | |||||  | ||||  |   |  |||  |  ||
Db     102366 TTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCCTGCCTGCTTGCCTGCTTGT 102425

Qy        689 TAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGC 748
              | |   | || || | ||||| ||  |  |||  |  ||||  |   | || || | ||||
Db     102426 TTGCTTGCCTGCTTGCCTGCTTGTTTGCTTGCCTGCTTGCTTGCTTGCCTGCTTGCCTGC 102485

Qy        749 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 808
                |   | |||  | ||||  |   | |||||  | |||||  |   | ||||  |  |||
Db     102486 CTGCCTGCTTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCCTGCCTGCTTGC 102545

Qy        809 TAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCT 845
                |   | |||||||||  |||||||||||||  ||| |
Db     102546 CTGCCTGCCTGCCTGCTTGCTTGCTTGCTTGGTTGTT 102582


RESULT 14
AC107913
LOCUS       AC107913               128953 bp    DNA     linear   PRI 28-AUG-2002
DEFINITION  Homo sapiens chromosome 17, clone CTD-3051C7, complete sequence.
ACCESSION   AC107913
VERSION     AC107913.7  GI:22532638
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 128953)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone CTD-3051C7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 128953)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 128953)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 128953)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 28, 2002 this sequence version replaced gi:22297439.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L23044
                Center clone name: 3051_C_7
            --------------
FEATURES                Location/Qualifiers
```

```
source          1. .128953
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="17"
                /map="17"
                /clone="CTD-3051C7"
                /clone_lib="CITD2 Human BAC"
repeat_region   386. .461
                /rpt_family="CT-rich"
repeat_region   complement(1888. .1975)
                /rpt_family="tRNA-Arg-AGA"
repeat_region   2505. .2586
                /rpt_family="tRNA-Leu-CTA"
repeat_region   complement(3077. .3148)
                /rpt_family="tRNA-Gln-CAG"
repeat_region   3246. .3523
                /rpt_family="MER33"
repeat_region   complement(3670. .3745)
                /rpt_family="tRNA-Lys-AAA"
repeat_region   5228. .5356
                /rpt_family="MIR3"
repeat_region   complement(5464. .5758)
                /rpt_family="AluSq"
repeat_region   5760. .5812
                /rpt_family="MER91"
repeat_region   complement(6843. .7005)
                /rpt_family="L2"
repeat_region   8462. .8650
                /rpt_family="C-rich"
repeat_region   complement(8858. .9155)
                /rpt_family="AluSp"
repeat_region   complement(9226. .9529)
                /rpt_family="AluSx"
repeat_region   complement(9582. .9876)
                /rpt_family="L2"
repeat_region   9965. .10118
                /rpt_family="FAM"
repeat_region   11131. .11308
                /rpt_family="MIR"
repeat_region   12520. .12540
                /rpt_family="GC_rich"
repeat_region   13116. .13238
                /rpt_family="C-rich"
repeat_region   complement(14519. .14584)
                /rpt_family="MIR"
repeat_region   14825. .15001
                /rpt_family="(TTTC)n"
repeat_region   15002. .15390
                /rpt_family="(CTTG)n"
repeat_region   complement(15391. .15680)
                /rpt_family="AluY"
repeat_region   complement(15684. .15965)
                /rpt_family="AluSq"
repeat_region   complement(16437. .16572)
                /rpt_family="AluSc"
repeat_region   complement(16656. .16785)
                /rpt_family="FLAM_A"
repeat_region   17009. .17100
                /rpt_family="L2"
repeat_region   complement(17263. .17542)
                /rpt_family="AluJo"
repeat_region   complement(17569. .18117)
                /rpt_family="L2"
repeat_region   complement(18137. .18354)
                /rpt_family="L2"
repeat_region   19136. .19250
                /rpt_family="(TTGGGG)n"
repeat_region   complement(20134. .20296)
                /rpt_family="L2"
repeat_region   complement(20313. .20372)
                /rpt_family="L2"
repeat_region   complement(20407. .20614)
                /rpt_family="L2"
repeat_region   complement(20717. .20993)
                /rpt_family="AluSg"
repeat_region   21268. .21571
                /rpt_family="AluJb"
repeat_region   21879. .22182
                /rpt_family="AluSq"
repeat_region   complement(22225. .22349)
                /rpt_family="MIR"
repeat_region   22422. .22595
                /rpt_family="FAM"
repeat_region   complement(22673. .22850)
                /rpt_family="MLT1H1"
repeat_region   complement(22902. .23026)
                /rpt_family="MLT1H1"
repeat_region   23051. .23106
                /rpt_family="CT-rich"
repeat_region   complement(23157. .23416)
                /rpt_family="AluJo"
repeat_region   complement(23521. .23824)
                /rpt_family="AluSx"
repeat_region   23863. .24015
                /rpt_family="MIR"
repeat_region   24050. .24358
                /rpt_family="AluJo"
repeat_region   24377. .24669
                /rpt_family="L1M4"
repeat_region   24670. .24981
                /rpt_family="AluSx"
repeat_region   24982. .25142
```

Query Match 20.8%; Score 208.6; DB 9; Length 128953;
Best Local Similarity 67.7%; Pred. No. 2.1e-24;
Matches 292; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

```
Qy      222 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 281
Db    14961 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 15020

Qy      282 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 341
Db    15021 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 15080

Qy      342 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 401
Db    15081 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 15140

Qy      402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461
Db    15141 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 15200

Qy      462 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 521
Db    15201 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 15260

Qy      522 TGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGG 581
Db    15261 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 15320

Qy      582 GGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCT 641
Db    15321 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 15380

Qy      642 GGGCTGCTAGT 652
Db    15381 TGCTTGCTTGT 15391
```

```
RESULT 15
AC098079
LOCUS       AC098079              276954 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-126O22, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
```

```
ACCESSION   AC098079
VERSION     AC098079.6  GI:30522704
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 276954)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 276954)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 276954)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:23269184.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GGFU
                Center clone name: CH230-126O22
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 219903 bases at least Q40
                Consensus quality: 222552 bases at least Q30
                Consensus quality: 224170 bases at least Q20
                Estimated insert size: 232969; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   271976: contig of 271976 bp in length
            *   271977   272076: gap of unknown length
            *   272077   276954: contig of 4878 bp in length.
FEATURES             Location/Qualifiers
     source          1. .276954
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-126O22"
     misc_feature    1. .1105
                     /note="wgs_contig"
     misc_feature    35576. .37460
                     /note="wgs_contig"
BASE COUNT    71863 a  46380 c  44551 g  62022 t  52138 others
ORIGIN

  Query Match             20.1%;  Score 201;  DB 2;  Length 276954;
  Best Local Similarity   57.9%;  Pred. No. 2.7e-23;
  Matches  357;  Conservative    0;  Mismatches  260;  Indels    0;  Gaps    0;

Qy        266 TGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTA 325
              || | | |   |||   | | ||||   ||   |  |||  ||| ||||   ||   |
Db     215110 TGGTGGTTGTTTGGTTGGTTGGCTGGTTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 215169

Qy        326 GGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTA 385
              |||  ||| ||| ||||||  ||| ||||   ||   || |    |||||| ||| |
Db     215170 GGCTGGCTGGCTGGCTGGCTGGCTGTTTGGTTGGTTGACTGGCTGGCTGGCTGGCT 215229

Qy        386 GGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 445
              ||   ||| ||||||  ||| ||| ||| ||||||  ||| ||| ||| ||||||  | |
Db     215230 GGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGTT 215289

Qy        446 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 505
              | | ||| | |||   | |  || ||| ||||||  ||| ||| ||| ||||||  |||
Db     215290 GGTTGGCTGGTTGGTTGGTTGACTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 215349
```

```
Qy        506 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGC 565
              ||| ||| |||||| ||| ||| ||| |||||| ||| | ||   ||  || ||
Db     215350 GGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGTTGGCTGGTTGGTTGGTT 215409

Qy        566 TGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGC 625
              ||| ||| |||    ||  ||   ||  ||  | ||| | | | || ||   | ||
Db     215410 GGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGTTGGTTGGCTGGTTGGTTGGTTGACT 215469

Qy        626 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGC 685
              ||| |  ||  |||||  ||| |  ||   |||||  ||| |  ||   |||||
Db     215470 GGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 215529

Qy        686 TGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGC 745
              ||| |  ||    |||    | | || |    |||||   ||| |  ||    ||||
Db     215530 GGCTGGCTGGCTGGCTGTTTGGTTGGTTGACTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 215589

Qy        746 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGC 805
              ||| |  ||   |||||   ||| |  ||   ||||    | | || |    |||||
Db     215590 GGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGTTTGGTTGGTTGACTGGCTGGCT 215649

Qy        806 TGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCT 865
              ||| |  ||   ||  ||||||| ||| ||| ||| ||| |  |    | | |   |||
Db     215650 GGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGTTTGGTTGGTTGGCT 215709

Qy        866 ACTAGGGCTGCTGTGCA 882
                |   ||  || | ||
Db     215710 GGTTTTGCCACTTTTCA 215726
```

```
RESULT 16
AC073686
LOCUS       AC073686               229010 bp    DNA     linear   HTG 29-JUN-2000
DEFINITION  Mus musculus clone RP23-137O3, WORKING DRAFT SEQUENCE, 24 unordered
            pieces.
ACCESSION   AC073686
VERSION     AC073686.1  GI:8810303
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 229010)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 229010)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 1789819
            Center clone name: RPCI-23_137O3
            -------
            Summary Statistics
            Consensus quality: 210402 bases at least Q40
            Consensus quality: 220249 bases at least Q30
            Consensus quality: 222247 bases at least Q20
            Estimated insert size: 231000; agarose-fp estimation
            Estimated insert size: 226710; sum-of-contigs estimation
            Quality coverage:  8.2 in Q20 bases; agarose-fp estimation
            Quality coverage:  8.35 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 24 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1213: contig of 1213 bp in length
            *     1214     1313: gap of unknown length
            *     1314     2782: contig of 1469 bp in length
            *     2783     2882: gap of unknown length
            *     2883     3987: contig of 1105 bp in length
            *     3988     4087: gap of unknown length
            *     4088     5464: contig of 1377 bp in length
            *     5465     5564: gap of unknown length
            *     5565     6708: contig of 1144 bp in length
            *     6709     6808: gap of unknown length
            *     6809     9727: contig of 2919 bp in length
            *     9728     9827: gap of unknown length
            *     9828    11170: contig of 1343 bp in length
            *    11171    11270: gap of unknown length
            *    11271    14915: contig of 3645 bp in length
            *    14916    15015: gap of unknown length
            *    15016    17525: contig of 2510 bp in length
            *    17526    17625: gap of unknown length
            *    17626    19558: contig of 1933 bp in length
            *    19559    19658: gap of unknown length
            *    19659    24103: contig of 4445 bp in length
            *    24104    24203: gap of unknown length
            *    24204    31169: contig of 6966 bp in length
            *    31170    31269: gap of unknown length
            *    31270    36961: contig of 5692 bp in length
            *    36962    37061: gap of unknown length
            *    37062    44441: contig of 7380 bp in length
            *    44442    44541: gap of unknown length
            *    44542    52366: contig of 7825 bp in length
            *    52367    52466: gap of unknown length
            *    52467    61496: contig of 9030 bp in length
            *    61497    61596: gap of unknown length
            *    61597    77798: contig of 16202 bp in length
            *    77799    77898: gap of unknown length
            *    77899    88928: contig of 11030 bp in length
            *    88929    89028: gap of unknown length
            *    89029   102867: contig of 13839 bp in length
            *   102868   102967: gap of unknown length
            *   102968   119649: contig of 16682 bp in length
            *   119650   119749: gap of unknown length
            *   119750   138547: contig of 18798 bp in length
            *   138548   138647: gap of unknown length
            *   138648   157565: contig of 18918 bp in length
            *   157566   157665: gap of unknown length
            *   157666   184818: contig of 27153 bp in length
            *   184819   184918: gap of unknown length
            *   184919   229010: contig of 44092 bp in length.
FEATURES             Location/Qualifiers
     source          1. .229010
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-137O3"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT    63071 a  50891 c  49844 g  62897 t   2307 others
ORIGIN
```

```
  Query Match           19.9%;  Score 199.4;  DB 2;  Length 229010;
  Best Local Similarity   62.9%;  Pred. No. 5.2e-23;
  Matches  325;  Conservative    0;  Mismatches  191;  Indels    1;  Gaps    1;

Qy        329 CTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGG 388
              ||| ||| | ||| |||||||||| |     | | |   |  ||| ||||| | |   |
Db      37960 CTGTTTGTTTGCTTGCCTGCTTGTTTGCTTGCCTGCCTGCTTGCCTGCTTGTTTGCTTGC 38019

Qy        389 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 448
              ||||| | | || ||  ||| |||  ||| |||||| || ||||| ||  || || ||||
Db      38020 CTGCTTGTTTGCTTGTCTGCCTGCCTGCTTGCCTGCCTGTTTGCTTGCCTGCTTGTTTGC 38079

Qy        449 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 508
```

```
Db      38080 TTGCCTGCTTGTTTGCTTGCCTGCTTGCTTGCCTGCTTGCTTGCCTGACTGCCTGCCTGC 38139

Qy        509 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGC 568

Db      38140 TTGCTTGCTTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGC 38199

Qy        569 TTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGC 628

Db      38200 TTGCCTGCTTGCCTGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGTTTGC 38259

Qy        629 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 688

Db      38260 TTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCCTGCCTGCTTGCCTGCTTGG 38319

Qy        689 TAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGC 748

Db      38320 TTGCTTGCCTGCTTGCCTGCTTGTTGCT-TGCCTGCTTGCTTGCTTGCCTGCTTGCCTGC 38378

Qy        749 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 808

Db      38379 CTGCCTGCTTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCCTGCCTGCTTGC 38438

Qy        809 TAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCT 845

Db      38439 CTGCCTGCCTGCCTGCTTGCTTGCTTGCTTGGTTGTT 38475
```

```
RESULT 17
AC117991
LOCUS       AC117991              271190 bp    DNA     linear   HTG 13-MAR-2003
DEFINITION  Mus musculus clone RP23-148J22, WORKING DRAFT SEQUENCE, 17
            unordered pieces.
ACCESSION   AC117991
VERSION     AC117991.3  GI:28933908
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 271190)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-148J22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 271190)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 271190)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 13, 2003 this sequence version replaced gi:28209782.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L15575
                Center clone name: 148_J_22
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 267377 bases at least Q40
                Consensus quality: 268701 bases at least Q30
                Consensus quality: 269200 bases at least Q20
                Insert size: 236000; agarose-fp
                Insert size: 269590; sum-of-contigs
                Quality coverage: 10.8 in Q20 bases; agarose-fp
                Quality coverage: 9.5 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      710: contig of 710 bp in length
            *      711      810: gap of 100 bp
            *      811     5531: contig of 4721 bp in length
            *     5532     5631: gap of 100 bp
            *     5632     7825: contig of 2194 bp in length
            *     7826     7925: gap of 100 bp
            *     7926    13529: contig of 5604 bp in length
            *    13530    13629: gap of 100 bp
            *    13630    19264: contig of 5635 bp in length
            *    19265    19364: gap of 100 bp
            *    19365    25381: contig of 6017 bp in length
            *    25382    25481: gap of 100 bp
            *    25482    38157: contig of 12676 bp in length
            *    38158    38257: gap of 100 bp
            *    38258    52292: contig of 14035 bp in length
            *    52293    52392: gap of 100 bp
            *    52393    70736: contig of 18344 bp in length
            *    70737    70836: gap of 100 bp
            *    70837    82153: contig of 11317 bp in length
            *    82154    82253: gap of 100 bp
```

```
           *    82254   101307: contig of 19054 bp in length
           *   101308   101407: gap of 100 bp
           *   101408   120674: contig of 19267 bp in length
           *   120675   120774: gap of 100 bp
           *   120775   149796: contig of 29022 bp in length
           *   149797   149896: gap of 100 bp
           *   149897   176329: contig of 26433 bp in length
           *   176330   176429: gap of 100 bp
           *   176430   202707: contig of 26278 bp in length
           *   202708   202807: gap of 100 bp
           *   202808   234746: contig of 31939 bp in length
           *   234747   234846: gap of 100 bp
           *   234847   271190: contig of 36344 bp in length.
FEATURES             Location/Qualifiers
     source          1. .271190
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-148J22"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .710
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    811. .5531
                     /note="assembly_fragment"
     misc_feature    5632. .7825
                     /note="assembly_fragment"
     misc_feature    7926. .13529
                     /note="assembly_fragment"
     misc_feature    13630. .19264
                     /note="assembly_fragment"
     misc_feature    19365. .25381
                     /note="assembly_fragment"
     misc_feature    25482. .38157
                     /note="assembly_fragment"
     misc_feature    38258. .52292
                     /note="assembly_fragment"
     misc_feature    52393. .70736
                     /note="assembly_fragment"
     misc_feature    70837. .82153
                     /note="assembly_fragment"
     misc_feature    82254. .101307
                     /note="assembly_fragment"
     misc_feature    101408. .120674
                     /note="assembly_fragment"
     misc_feature    120775. .149796
                     /note="assembly_fragment"
     misc_feature    149897. .176329
                     /note="assembly_fragment"
     misc_feature    176430. .202707
                     /note="assembly_fragment"
     misc_feature    202808. .234746
                     /note="assembly_fragment"
     misc_feature    234847. .271190
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    70565 a  63542 c  62426 g  73055 t   1602 others
ORIGIN

  Query Match             19.8%;  Score 197.8;  DB 2;  Length 271190;
  Best Local Similarity   57.1%;  Pred. No. 8.9e-23;
  Matches  380;  Conservative    0;  Mismatches  282;  Indels    3;  Gaps    1;

Qy        214 GGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 273
              |||  |   ||  ||  ||| |||| |   || ||||   ||   || ||   ||| |||
Db     186704 GGTTGGTTGGCTGGCTGGCTGGCTGTCAGTCTGGCTGGTTGGCTGGTTGGTTAGCTGGCT 186763

Qy        274 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCT 333
              ||||||  ||| ||||   ||  ||| ||   | | | ||   ||   || ||   | |
Db     186764 GGCTGGCTGGCTGGCTGGTTGGTTAGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 186823

Qy        334 TGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCT 393
              ||| |||||   ||| ||||   ||   |  ||    | | ||  || || ||   |||
Db     186824 GGCTGGCTGGTTGGCTGGCTGGTTGGCTGGCTGGCTGTCAGTCTGTCTGGTTGGTTGGCT 186883

Qy        394 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 453
              ||||||  ||| ||| ||| ||||||  ||| ||| ||| |||||   ||| | | | |
Db     186884 GGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGTTGGCTGGTTGGTT 186943

Qy        454 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 513
              ||| |   | | ||| ||| ||||||  ||| | | ||| | ||||  | | | | | |
Db     186944 GGCTAGTTAGTTGGCTGGCTGGCTGGCTGGCTGGTTGGCTGGTTGGCTGGTTGGTTGGTT 187003

Qy        514 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCT 573
              | ||||  ||| ||| |   ||| |  | | |||   ||  ||  ||  ||| ||| |||
Db     187004 GGTTGGCTGGCTGGCTGTCAGTCTGTCTGGTTGGCTGGTTGGTTAGCTGGCTGGCTGGCT 187063

Qy        574 TGCTAAGGGGACGGT---ACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTA 630
              |||    ||   ||     |  ||  || | | ||| | |   || ||   |||||
Db     187064 GGCTGGTTGGTTAGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGGCT 187123

Qy        631 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 690
              |   ||||| | | |||   |   ||||| | |||||   |   |||||   |||||
Db     187124 GGCTGGCTGTTAGTCTGTCTGGTTGGCTGGTTGGCTGGCTGGCTGGCTGGCTGGCT 187183

Qy        691 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTA 750
              |   ||||    ||| | | ||  | ||| | ||||| | ||  || |  | || ||
Db     187184 GGCTGGCTGGCTGGCAGGTTGTCAGTCTGGTTGGCTGGTTGTTTGGGTGGTTGGTTGGCT 187243

Qy        751 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 810
              |   |||||   |||| | |   || || | || || | |   || || | || || |
Db     187244 GACTGGCTGGCTGGCTGGTTGGTTGGTTGTTTGGGTGGTTGGTTGGTTGGTTGGTTGGTT 187303

Qy        811 GTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAG 870
              |   ||||| |||   | ||| ||| ||| ||| | |   ||  | ||| |||
Db     187304 GGCTGGCTGGCTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCTAGCT 187363

Qy        871 GGCTG 875
              |||||
Db     187364 GGCTG 187368


RESULT 18
AC132225
LOCUS       AC132225              219713 bp    DNA     linear   HTG 03-SEP-2002
DEFINITION  Mus musculus chromosome UNK clone RP24-558H21, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC132225
VERSION     AC132225.1  GI:22657606
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 219713)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 219713)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-SEP-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT
            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------- Project Information ----------
            Center project name: M_BB0558H21
            ----------
            ---------- Summary Statistics ----------
            Sequencing vector: M13; 0%
```

```
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 216515 bases at least Q40
            Consensus quality: 216969 bases at least Q30
            Consensus quality: 217277 bases at least Q20
            Insert size: 208000; agarose-fp
            Insert size: 218913; sum-of-contigs
            Quality coverage: 14.28 in Q20 bases; agarose-fp
            Quality coverage: 10.95 in Q20 bases; sum-of-contigs
            ----------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    11635: contig of 11635 bp in length
            *    11636    11735: gap of unknown length
            *    11736    24748: contig of 13013 bp in length
            *    24749    24848: gap of unknown length
            *    24849    41932: contig of 17084 bp in length
            *    41933    42032: gap of unknown length
            *    42033    66899: contig of 24867 bp in length
            *    66900    66999: gap of unknown length
            *    67000    95249: contig of 28250 bp in length
            *    95250    95349: gap of unknown length
            *    95350   143163: contig of 47814 bp in length
            *   143164   143263: gap of unknown length
            *   143264   202040: contig of 58777 bp in length
            *   202041   202140: gap of unknown length
            *   202141   209252: contig of 7112 bp in length
            *   209253   209352: gap of unknown length
            *   209353   219713: contig of 10361 bp in length.
FEATURES             Location/Qualifiers
     source          1. .219713
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP24-558H21"
     misc_feature    1. .11635
                     /note="assembly_name:Contig10"
     misc_feature    11736. .24748
                     /note="assembly_name:Contig11"
     misc_feature    24849. .41932
                     /note="assembly_name:Contig12"
     misc_feature    42033. .66899
                     /note="assembly_name:Contig13"
     misc_feature    67000. .95249
                     /note="assembly_name:Contig14"
     misc_feature    95350. .143163
                     /note="assembly_name:Contig15"
     misc_feature    143264. .202040
                     /note="assembly_name:Contig16"
     misc_feature    202141. .209252
                     /note="assembly_name:Contig8"
     misc_feature    209353. .219713
                     /note="assembly_name:Contig9"
BASE COUNT    58319 a  48550 c  50247 g  61747 t    850 others
ORIGIN

  Query Match             19.7%;  Score 196.8;  DB 2;  Length 219713;
  Best Local Similarity   55.6%;  Pred. No. 1.3e-22;
  Matches  378;  Conservative    0;  Mismatches  302;  Indels    0;  Gaps    0;

Qy        185 GACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGC 244
              | | | | | |   ||  |  || || | ||  | ||||||||||    ||  ||||||
Db      82403 GTCCTCATCTTCTATCTCCTTTGAGACAGAGTCTGCCTGCCTGCCTATCTGCCTGCCTGC 82462

Qy        245 TTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGC 304
              |||      | | |   |||||| |||      ||||||  |||      | |  |  ||
Db      82463 CTGCCTCTCTGCCTGCCTGCCTGCCTGCCTCTCTGCCTGCCTGCCTGCCTGCCTCTCTGC 82522

Qy        305 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 364
              ||||  |||        | |   |||||| |||  ||  ||||||  |||        | |
Db      82523 CTGCCTGCCTGCATATCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTATCTGC 82582

Qy        365 AGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 424
              | ||||  ||  |||   |  | ||||  ||  |||||| ||| |||  ||    ||||
Db      82583 CTGCCTGCCTGCATGCCTATCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCTCTGC 82642

Qy        425 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 484
              |||  |||       ||||||  |||  |||  ||  ||||||  |||  |||  ||  ||||||
Db      82643 CTGCCTGCCTCTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGC 82702

Qy        485 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 544
              |   |||  ||  ||||||  |||  |    ||  ||  |||  |||  |||      ||||||
Db      82703 CTATCTGCCTGCCTGCCTGCCTGCCTCTCTGCCTGCTTGCCTGCCTGCCTCTCTGCCTGC 82762

Qy        545 TTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGC 604
              |||        | |   | |||| |||  |||  |   |  | | |   |||||||||   ||
Db      82763 CTGCCTGCCTATCTGCCTGCCTGCCTGCCTGCCTATCTGCCTGCCTGCCTGCCTGCCTGC 82822

Qy        605 TTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 664
              ||    |  | ||||  | ||||  |   | ||||    ||||  |   | ||||
Db      82823 CTGCCTATCTGCCTGCCTGCCTGCCTGCCTGCCTCTCTGCCTGCCTGCCTGCCTAT 82882

Qy        665 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGG 724
              ||||  |   | ||||  | ||||  |   | |||   | ||||  |   | ||||  |
Db      82883 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCTCTGCCTGCCTGCCTGCCTGCCTGC 82942

Qy        725 CTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 784
              ||||  |   | ||  |  | ||||  |   | ||||  | ||||  |   | ||||  |
Db      82943 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTATCTGCCTGCCTGC 83002

Qy        785 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGC 844
              ||||  ||  | ||||  | ||||  |   | ||||||||  ||  |||  |   |||  |||
Db      83003 CTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTATCTGCCTGC 83062

Qy        845 TAGTGGGGCCGCTTGCCTGC 864
              |     | ||  |||  |||
Db      83063 CTGCCTCTCTGCCTGCTTGC 83082


RESULT 19
AP000291
LOCUS       AP000291               27213 bp    DNA     linear   PRI 20-NOV-1999
DEFINITION  Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region,
            clone:Q64C8, complete sequence.
ACCESSION   AP000291
VERSION     AP000291.1  GI:4835660
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 27213)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE     Homo sapiens 27,213bp genomic DNA of 21q22.1
  JOURNAL   Published Only in DataBase (1999)
REFERENCE   2  (bases 1 to 27213)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     The sequence is a part of the data (ACCESSION No. AP000174 -
```

```
          AP000194).
          The sequencing project is supported by Japan Science Technology
          Corporation (JST) and The Institute of Physical and Chemical
          Research (RIKEN).
FEATURES             Location/Qualifiers
     source          1. .27213
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="Q64C8"
BASE COUNT     7998 a   5908 c   5753 g   7554 t
ORIGIN

  Query Match             19.5%;  Score 195;  DB 9;  Length 27213;
  Best Local Similarity   58.8%;  Pred. No. 4.5e-22;
  Matches  336;  Conservative    0;  Mismatches  235;  Indels    0;  Gaps    0;

Qy        295 GACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 354
              ||| |   || || ||| | |       |   || || ||| | ||| |  |||||| |
Db      20357 GACTGGTTGCTTGGTTGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTT 20416

Qy        355 AGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTC 414
               |   |   || || | ||||| | |   || || | | | |  ||||||||||||
Db      20417 GGCTGGTTGCTTGGTTGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTT 20476

Qy        415 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 474
              ||| || ||||||||||| ||| || ||  || ||| | | | |  |||||| ||| |
Db      20477 GCTTGCTTGCTTGCTTGCTTGCTTGGCTGGTTGGTTGGTTGGTTGCTTGGTTGGTT 20536

Qy        475 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 534
              | | || |||||||||| | ||| || |  ||| ||||| | | || ||||  ||| |
Db      20537 GGTTGCTTGCTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCATTGGTT 20596

Qy        535 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCT 594
              ||| |  ||||||||  |  |   |   || || |||||||  |  |  |   |  | |
Db      20597 GCTTGGTTGCTTGCTTCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTT 20656

Qy        595 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 654
              |  || | | ||| | |   || |||| || |||| |   || || | || || | ||
Db      20657 GGTTGGTTGGTTGCTTGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGTTT 20716

Qy        655 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 714
              || |||| || || | |   || || | |  || | |   || |  | |  |||| |
Db      20717 GGTTGCTTGGTTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTT 20776

Qy        715 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 774
               | |||| |  || |     || || | || || | |   || || | || || | |
Db      20777 TGTTGCTTGCTTGGTTCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 20836

Qy        775 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTT 834
              || || |  | || | |   || || | || |||| |   || || ||| |   ||| ||
Db      20837 GGTTGGTTCGTTGATTGCTTGGTTGGTTGGTTGCTTGGTTGGTTGGCTGGTTCATTGGTT 20896

Qy        835 GCTTGCTTGCTAGTGGGGCCGCTTGCCTGCT 865
              | ||| ||||| |   |   | |||  || |
Db      20897 GGTTGGTTGCTTGGTTGTTTGGTTGGTTGGT 20927


RESULT 20
AP000043
LOCUS       AP000043              100000 bp    DNA     linear   PRI 20-NOV-1999
DEFINITION  Homo sapiens genomic DNA, chromosome 21q22.1, segment 14/28,
            complete sequence.
ACCESSION   AP000043
VERSION     AP000043.1  GI:3132353
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA, chromosome 21q
  JOURNAL   Published Only in DataBase (1998)
REFERENCE   2  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAY-1998) Masahira Hattori, Kitasato University,
            Department of Science, JST Sequencing Laboratory; Kitasato 1-15-1,
            Sagamihara 228, Japan (E-mail:hattori@hgc.ims.u-tokyo.ac.jp,
            Tel:0427-78-9732, Fax:0427-78-9561)
COMMENT     This sequence is conducted by Kitasato University JST sequencing
            Laboratory as a JST sequencing team.
            Principal Investigator:Yoshiyuki Sakaki Ph.D.
            Phone: +81-3-5449-5622, Fax : +81-3-5449-5445,
            sakaki@hgc.ims.u-tokyo.ac.jp
            Sub-leader: Tadayoshi Shiba Ph.D., Masahira Hattori Ph.D. The
            sequence is submitted by:Human Genome Sequencing in ALIS project of
            JST
            Japan Science and Technology Corporation (JST)
            5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0028 Japan
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www-alis.tokyo.jst.go.jp/HGS/top. html)
            or send email to webmaster@www-alis.tokyo.jst.go.jp'.
FEATURES             Location/Qualifiers
     source          1. .100000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="T1866-f90A8f_7"
BASE COUNT     27784 a  22015 c  21661 g  28540 t
ORIGIN

  Query Match             19.5%;  Score 195;  DB 9;  Length 100000;
  Best Local Similarity   58.8%;  Pred. No. 3.2e-22;
  Matches  336;  Conservative    0;  Mismatches  235;  Indels    0;  Gaps    0;

Qy        295 GACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 354
              ||| |   || || ||| | |       |   || || ||| | ||| |  |||||| |
Db      12941 GACTGGTTGCTTGGTTGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTT 13000

Qy        355 AGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTC 414
               |   |   || || | ||||| | |   || || | | | |  ||||||||||||
Db      13001 GGCTGGTTGCTTGGTTGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTT 13060

Qy        415 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 474
              ||| || ||||||||||| ||| || ||  || ||| | | | |  |||||| ||| |
Db      13061 GCTTGCTTGCTTGCTTGCTTGCTTGGCTGGTTGGTTGGTTGGTTGCTTGGTTGGTT 13120

Qy        475 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 534
              | | || |||||||||| | ||| || |  ||| ||||| | | || ||||  ||| |
Db      13121 GGTTGCTTGCTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCATTGGTT 13180

Qy        535 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCT 594
              ||| |  ||||||||  |  |   |   || || |||||||  |  |  |   |  | |
Db      13181 GCTTGGTTGCTTGCTTCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTT 13240

Qy        595 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 654
              |  || | | ||| | |   || |||| || |||| |   || || | || || | ||
Db      13241 GGTTGGTTGGTTGCTTGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGTTT 13300

Qy        655 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 714
              || |||| || || | |   || || | |  || | |   || |  | |  |||| |
Db      13301 GGTTGCTTGGTTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTT 13360

Qy        715 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 774
               | |||| |  || |     || || | || || | |   || || | || || | |
Db      13361 TGTTGCTTGCTTGGTTCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 13420

Qy        775 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTT 834
              || || |  | || | |   || || | || |||| |   || || ||| |   ||| ||
```

```
Db      13421 GGTTGGTTCGTTGATTGCTTGGTTGGTTGGTTGCTTGGTTGGTTGGCTGGTTCATTGGTT 13480

Qy        835 GCTTGCTTGCTAGTGGGGCCGCTTGCCTGCT 865
              | ||| ||||| |   |   | |||  || |
Db      13481 GGTTGGTTGCTTGGTTGTTTGGTTGGTTGGT 13511


RESULT 21
AP000111
LOCUS       AP000111              100000 bp    DNA     linear   PRI 25-MAY-2002
DEFINITION  Homo sapiens genomic DNA of 21q22.1, GART and AML related,
            Q78C10-149C3 region, segment 14/20.
ACCESSION   AP000111
VERSION     AP000111.1  GI:4730845
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Homo sapiens 2,051,516bp genomic DNA of 21q22.1 GART and AML region
  JOURNAL   Published Only in Database (1999)
REFERENCE   2  (bases 1 to 100000)
  AUTHORS   Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-1999) Mika Hirakawa, Japan Science and Technology
            Corporation (JST), Advanced Databases Department; 5-3, Yonbancho,
            Chiyoda-ku, Tokyo 102-0081, Japan (E-mail:mika@tokyo.jst.go.jp,
            URL:http://www-alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
            Fax:81-3-5214-8470)
COMMENT     This sequence is conducted by Kitasato University JST sequencing
            Laboratory as a JST sequencing team.
            Principal Investigator:Yoshiyuki Sakaki Ph.D.
            Phone: +81-3-5449-5622, Fax : +81-3-5449-5445,
            sakaki@hgc.ims.u-tokyo.ac.jp
            Sub-leader: Tadayoshi Shiba Ph.D., Masahira Hattori Ph.D. The
            sequence is submitted by Human Genome Sequencing in ALIS project of
            JST.
            Japan Science and Technology Corporation (JST)
            5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081 Japan
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www-alis.tokyo.jst.go.jp/HGS/) or send
            email to webmaster@www-alis.tokyo.jst.go.jp.
FEATURES             Location/Qualifiers
     source          1. .100000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
     STS             66264. .66403
                     /note="A006D32;The location is between each flanking site
                     of PCR primers."
                     /db_xref="GDB:4571163"
     STS             66282. .66515
                     /note="SHGC-52112;The location is between each flanking
                     site of PCR primers."
                     /db_xref="GDB:6464755"
     STS             92086. .92206
                     /note="SHGC-52139;The location is between each flanking
                     site of PCR primers."
                     /db_xref="GDB:6464581"
     STS             92099. .92279
                     /note="RH18027;The location is between each flanking site
                     of PCR primers."
                     /db_xref="GDB:4572813"
BASE COUNT    27992 a  22056 c  21621 g  28331 t
ORIGIN

  Query Match           19.5%;  Score 195;  DB 9;  Length 100000;
  Best Local,Similarity 58.8%;  Pred. No. 3.2e-22;

  Matches  336;  Conservative    0;  Mismatches  235;  Indels    0;  Gaps    0;

Qy        295 GACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 354
              ||| |   || || ||| |  |       |   || || ||| | ||| |  |||||| |
Db      18917 GACTGGTTGCTTGGTTGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTT 18976

Qy        355 AGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTC 414
              |  |   |   || || | ||||| | |   || || | | | | ||||||||||||
Db      18977 GGCTGGTTGCTTGGTTGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTT 19036

Qy        415 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 474
              ||| || |||||||||||| ||| || ||  || ||| | | | |  |||||| ||| |
Db      19037 GCTTGCTTGCTTGCTTGCTTGCTTGCTTGGCTGGTTGGTTGGTTGGTTGCTTGGTTGGTT 19096

Qy        475 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 534
              | | || |||||||||| | ||| || |  ||| |||||  | | || |||||  ||| |
Db      19097 GGTTGCTTGCTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCATTGGTT 19156

Qy        535 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCT 594
              ||| |  |||||||  |  |   |   || || ||||||| |  |  |   |   |  |
Db      19157 GCTTGGTTGCTTGCTTCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTT 19216

Qy        595 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 654
              |  || | | ||| | |   || |||| || |||| |   || || | || || | ||
Db      19217 GGTTGGTTGGTTGCTTGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGTTT 19276

Qy        655 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 714
              || |||| || || | |   || || | |  || | |   || |  | |  |||| |
Db      19277 GGTTGCTTGGTTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTT 19336

Qy        715 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 774
              | |||| |  || |     || || | || || | |   || || | || || | | | |
Db      19337 TGTTGCTTGCTTGGTTCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 19396

Qy        775 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTT 834
              || || |  | || | |   || || | || |||| |   || || ||| |   ||| ||
Db      19397 GGTTGGTTCGTTGATTGCTTGGTTGGTTGGTTGCTTGGTTGGTTGGCTGGTTCATTGGTT 19456

Qy        835 GCTTGCTTGCTAGTGGGGCCGCTTGCCTGCT 865
              | ||| ||||| |   |   | |||  || |
Db      19457 GGTTGGTTGCTTGGTTGTTTGGTTGGTTGGT 19487


RESULT 22
AP000187
LOCUS       AP000187              100000 bp    DNA     linear   PRI 08-JAN-2000
DEFINITION  Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region,
            clone Q78C10-f32E9, segment 14/21, complete sequence.
ACCESSION   AP000187
VERSION     AP000187.1  GI:4827086
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE     Homo sapiens 2,083,744bp genomic DNA of 21q22.1 (REGION:
            D21S226-AML CLONE RANGE: Q78C10-f32E9)
  JOURNAL   Published Only in DataBase (1999)
REFERENCE   2  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     E. coli transposon insertion:The present data does not contain E.
            coli transposon sequences which integrated in the
```

```
          original/previous sequences.  We determined the boundary between
          the insertion and genomic sequences experimentally, removed the
          insertion sequences, reconstituted the present data. The sequencing
          project is supported by Japan Science Technology Corporation (JST)
          and The Institute of Physical and Chemical Research (RIKEN).
FEATURES             Location/Qualifiers
     source          1. .100000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
BASE COUNT    27992 a  22056 c  21621 g  28331 t
ORIGIN

  Query Match             19.5%;  Score 195;  DB 9;  Length 100000;
  Best Local Similarity   58.8%;  Pred. No. 3.2e-22;
  Matches  336;  Conservative    0;  Mismatches  235;  Indels    0;  Gaps    0;

Qy        295 GACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 354
              ||| |   || || ||| |  |      |   || || ||| | ||| |  |||||| |
Db      18917 GACTGGTTGCTTGGTTGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTT 18976

Qy        355 AGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTC 414
              |  |   |   || || | ||||| | |   || || | | | |  ||||||||||||
Db      18977 GGCTGGTTGCTTGGTTGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTT 19036

Qy        415 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 474
              ||| || |||||||||||| ||| || ||  || ||| | | | |  |||||| ||| |
Db      19037 GCTTGCTTGCTTGCTTGCTTGCTTGGCTGGTTGGTTGGTTGGTTGCTTGGTTGGTT 19096

Qy        475 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 534
              | | || |||||||||| | ||| || |  ||| ||||| | | || |||||  ||| |
Db      19097 GGTTGCTTGCTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCATTGGTT 19156

Qy        535 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCT 594
              ||| |  |||||||  |  |   |   || || |||||| |  |  |   |   |  |
Db      19157 GCTTGGTTGCTTGCTTCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTT 19216

Qy        595 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 654
              |  || | | ||| | |   || |||| || |||| |   || || | || || | ||
Db      19217 GGTTGGTTGGTTGCTTGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGTTT 19276

Qy        655 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 714
              || |||| || || | |   || || | |  || | |   || |  | |  |||| |
Db      19277 GGTTGCTTGGTTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTT 19336

Qy        715 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 774
              | |||| |  || |    || || | || || | |   || || | || || | |
Db      19337 TGTTGCTTGCTTGGTTCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 19396

Qy        775 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTT 834
              || || |  | || | |   || || | || |||| |   || || ||| |   ||| ||
Db      19397 GGTTGGTTCGTTGATTGCTTGGTTGGTTGGTTGCTTGGTTGGTTGGCTGGTTCATTGGTT 19456

Qy        835 GCTTGCTTGCTAGTGGGGCCGCTTGCCTGCT 865
              | ||| ||||| |   |   | |||  || |
Db      19457 GGTTGGTTGCTTGGTTGTTTGGTTGGTTGGT 19487


RESULT 23
AP001716
LOCUS       AP001716               340000 bp    DNA     linear   PRI 21-MAY-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 21q, section 60/105.
ACCESSION   AP001716 AL163261 BA000005
VERSION     AP001716.1  GI:7768717
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
            Asakawa,S., Shintani,A., Sasaki,K., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
            Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
            Hennig,S., Riesselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K.,
            Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
            Yaspo,M.L.
  TITLE     The DNA sequence of human chromosome 21
  JOURNAL   Nature 405 (6784), 311-319 (2000)
  MEDLINE   20289799
   PUBMED   10830953
REFERENCE   2  (bases 1 to 340000)
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
            Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
            Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
            Hennig,S., Riesselmann,L., Dagand,E., Wehrmaeyer,S., Borzym,K.,
            Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
            Yaspo,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing
            Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
            Group * Institute of Molecular Biotechnology, Genome Analysis *
            Keio University School of Medicine, Dept. of Molecular Biology *
            GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
            Genetics (addresses see below)
COMMENT     On May 30, 2000 this sequence version replaced gi:7717330.
            The chromosome 21 mapping and sequencing consortium consisting of
            * RIKEN Genomic Sciences Center, Human Genome Research Group, *
            Sagamihara 228-8555, Japan,
            * e.mail: hattori@gsc.riken.go.jp
            * URL: http://hgp.gsc.riken.go.jp/
            and
            * Institute of Molecular Biotechnology, Genome Analysis, *
            Beutenbergstrasse 11, D-07745 Jena, Germany,
            * e.mail: gscj-submit@genome.imb-jena.de
            * URL: http://genome.imb-jena.de/
            and
            * Keio University School of Medicine, Molecular Biology, * Tokyo
            160-8582, Japan,
            * e.mail: nshimizu@dmb-med.keio.ac.jp
            * URL: http://www.dmb.med.keio.ac.jp/
            and
            * GBF, Dept. of Genome Analysis,
            * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
            info.genome@gbf.de
            * URL: http://genome.gbf.de/
            and
            * Max-Planck Institute for Molecular Genetics,
            * Ihnestrasse 73, D-14195 Berlin, Germany,
            * e.mail: info-chr21@molgen.mpg.de
            * URL: http://chr21.rz-berlin.mpg.de/
            AL163261: Submitted (10-Apr-2000).
FEATURES             Location/Qualifiers
     source          1. .340000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
     source          <1. .16929
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
```

```
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="pQ57C2, 5' partial"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000288"
     source          16360. .103735
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="pS701"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000289"
     source          103356. .120643
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="pQ14E2"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000290"
     source          118735. .145947
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="pQ64C8"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000291"
     source          144738. .182946
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="pQ95D4"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000292"
     source          182189. .189406
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="pQ20D3"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000293"
     source          189053. .205485
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="pD5C10"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000294"
     source          204813. .260943
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="pS318"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000295"
     source          260219. .278406
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="pD71A4"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000296"
     source          277776. .300712
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="pD5C11"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000297"
     source          300164. .325723
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="PQ38G8"
                     /clone_lib="RP1 PAC library"
                     /note="Accession No. AP000298"
     source          322992. .>340000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="PQ102G11, 3' partial"
                     /clone_lib="RP1 PAC library"
                     /note="Accession No. AP000299"
     repeat_region   complement(722. .1108)
                     /note="MLT1B"
                     /rpt_family="LTR/MaLR"
                     /rpt_type=DISPERSED
     repeat_region   complement(1428. .1508)
                     /note="MIR"
                     /rpt_family="SINE/MIR"
                     /rpt_type=DISPERSED
     repeat_region   1794. .1944
                     /note="MIR"
                     /rpt_family="SINE/MIR"

  Query Match             19.5%;  Score 195;  DB 9;  Length 340000;
  Best Local Similarity   58.8%;  Pred. No. 2.3e-22;
  Matches  336;  Conservative    0;  Mismatches  235;  Indels    0;  Gaps    0;

Qy        295 GACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 354
              ||| |   || || ||| |  |       |   || || ||| | ||| |  |||||| |
Db     139091 GACTGGTTGCTTGGTTGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTT 139150

Qy        355 AGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTC 414
               |  |   |   || || | ||||| | |   || || | | | |  ||||||||||||
Db     139151 GGCTGGTTGCTTGGTTGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTT 139210

Qy        415 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 474
              ||| || ||||||||||||| ||| || ||  || ||| | | | |  |||||| ||| |
Db     139211 GCTTGCTTGCTTGCTTGCTTGCTTGGCTGGTTGGTTGGTTGGTTGCTTGGTTGGTT 139270

Qy        475 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 534
              | | || |||||||||| | ||| || |  ||| ||||| | | || |||||  ||| |
Db     139271 GGTTGCTTGCTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCATTGGTT 139330

Qy        535 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCT 594
              ||| |  |||||||  |  |   |   || || |||||| |  |  |  |   |  |
Db     139331 GCTTGGTTGCTTGCTTCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTT 139390

Qy        595 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 654
              |  || | | ||| | |   || |||| || |||| |   || || | || || | ||
Db     139391 GGTTGGTTGGTTGCTTGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGTTT 139450

Qy        655 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 714
```

```
Db     139451 GGTTGCTTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTT 139510

Qy        715 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 774

Db     139511 TGTTGCTTGCTTGGTTCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 139570

Qy        775 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTT 834

Db     139571 GGTTGGTTCGTTGATTGCTTGGTTGGTTGGTTGCTTGGTTGGTTGGCTGGTTCATTGGTT 139630

Qy        835 GCTTGCTTGCTAGTGGGGCCGCTTGCCTGCT 865

Db     139631 GGTTGGTTGCTTGGTTGTTTGGTTGGTTGGT 139661
```

RESULT 24
U93872/c
LOCUS U93872 133661 bp DNA linear VRL 09-JUL-2001
DEFINITION Kaposi's sarcoma-associated herpesvirus glycoprotein M, DNA replication protein, glycoprotein, DNA replication protein, FLICE inhibitory protein and v-cyclin genes, complete cds, and tegument protein gene, partial cds.
ACCESSION U93872
VERSION U93872.2 GI:14627174
KEYWORDS .
SOURCE Human herpesvirus 8 (Kaposi's sarcoma-associated herpesvirus - Human herpesvirus 8)
ORGANISM Human herpesvirus 8
Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae; Rhadinovirus.
REFERENCE 1 (bases 29032 to 30108; 117733 to 118431)
AUTHORS Zhong,W., Wang,H., Herndier,B. and Ganem,D.
TITLE Restricted expression of Kaposi sarcoma-associated herpesvirus (human herpesvirus 8) genes in Kaposi sarcoma
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (13), 6641-6646 (1996)
MEDLINE 96270595
PUBMED 8692871
REFERENCE 2 (bases 17242 to 17856)
AUTHORS Neipel,F., Albrecht,J.C., Ensser,A., Huang,Y.Q., Li,J.J., Friedman-Kien,A.E. and Fleckenstein,B.
TITLE Human herpesvirus 8 encodes a homolog of interleukin-6
JOURNAL J. Virol. 71 (1), 839-842 (1997)
MEDLINE 97138401
PUBMED 8985427
REFERENCE 3 (bases 123309 to 124082)
AUTHORS Li,M., Lee,H., Yoon,D.W., Albrecht,J.C., Fleckenstein,B., Neipel,F. and Jung,J.U.
TITLE Kaposi's sarcoma-associated herpesvirus encodes a functional cyclin
JOURNAL J. Virol. 71 (3), 1984-1991 (1997)
MEDLINE 97184528
PUBMED 9032330
REFERENCE 4 (bases 122660 to 123226)
AUTHORS Thome,M., Schneider,P., Hofmann,K., Fickenscher,H., Meinl,E., Neipel,F., Mattmann,C., Burns,K., Bodmer,J.L., Schroter,M., Scaffidi,C., Krammer,P.H., Peter,M.E. and Tschopp,J.
TITLE Viral FLICE-inhibitory proteins (FLIPs) prevent apoptosis induced by death receptors
JOURNAL Nature 386 (6624), 517-521 (1997)
MEDLINE 97242415
PUBMED 9087414
REFERENCE 5 (bases 1 to 133661)
AUTHORS Neipel,F., Albrecht,J.C. and Fleckenstein,B.
TITLE Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus human herpesvirus 8: determinants of its pathogenicity?
JOURNAL J. Virol. 71 (6), 4187-4192 (1997)
MEDLINE 97296220
PUBMED 9151804
REFERENCE 6 (bases 1 to 133661)
AUTHORS Neipel,F., Albrecht,J.-C., Ensser,A., Huang,Y.-Q., Li,J.J., Friedman-Kien,A.E. and Fleckenstein,B.
TITLE The genome of human herpesvirus 8 cloned from Kaposi's sarcoma
JOURNAL Unpublished
REFERENCE 7 (bases 1 to 133661)
AUTHORS Neipel,F., Albrecht,J.-C., Ensser,A., Huang,Y.-Q., Li,J.J., Friedman-Kien,A.E. and Fleckenstein,B.
TITLE Direct Submission
JOURNAL Submitted (17-MAR-1997) Virology, University of Erlangen, Schlossgarten 4, Institut for Klinische und Molekulare, Erlangen 91054, Germany
REFERENCE 8 (bases 1 to 133661)
AUTHORS Neipel,F., Albrecht,J.-C., Ensser,A., Huang,Y.-Q., Li,J.J., Friedman-Kien,A.E. and Fleckenstein,B.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2001) Virology, University of Erlangen, Schlossgarten 4, Institut for Klinische und Molekulare, Erlangen 91054, Germany
REMARK Sequence update by submitter
COMMENT On Jul 9, 2001 this sequence version replaced gi:2246466.
FEATURES Location/Qualifiers
source 1. .133661
/organism="Human herpesvirus 8"
/mol_type="genomic DNA"
/db_xref="taxon:37296"
/note="sequence derived from overlapping Lambda phage clones isolated from two libraries of Sau3A partially digested DNA from Kaposi's sarcoma biopsies"
CDS 105. .959
/note="ORF K1"
/codon_start=1
/protein_id="AAB62619.1"
/db_xref="GI:2246494"
/translation="MFLYVVCSLAVCFRGLLSLSLQSSPNLCPGVISTPYTLTCPSNT
SLPTSWYCNDTRLLRLTQQTFTVDTLICNFSCVGQSGHRHSLWITWYPQPVLQTLCGQ
PSNTVTCGQHVTLYCSTSGNNVTVWHLPNGRNETVSQTKYYNFTLMNQTEGCYACSNG
LSSRLSNRLCFSARCANLTPETHTVSVSSTTGFRTFSTNRRGTTLDVLVMKEAKSTNL
HIQVHFLVFMTLVALIGTMCGILGTIIFAHCQKQRDSNKTVPQQLQDYYSLHDLCTED
YTQPVDWY"
CDS 1127. .2779
/note="ORF 04, complement control protein homolog"
/codon_start=1
/protein_id="AAB62602.1"
/db_xref="GI:2246477"
/translation="MAFLRQTLWTLWTFTMVIGQDNEKCSQKTLIGYRLKMSRDGDIA
VGETVELRCRSGYTTYARNITATCLQGGTWSEPTATCNKKSCPNPGEIQNGKVIFHGG
QDALKYGANISYVCNEGYFLVGREYVRYCMIGASGQMAWSSSPPFCEKEKCHRPKIEN
GDFKPDKDYYEYNDAVHFECNEGYTLVGPHSIACAVNNTWTSNMPTCELAGCKFPSVT
HGYPIQGFSLTYKHKQSVTFACNDGFVLRGSPTITCNVTEWDPPLPKCVLEDIDDPNN
SNPGRLHPTPNEKPNGNVFQRSNYTEPPTKPEDTHTAATCDTNCEQPPKILPTSEGFN
ETTTSNTITKQLEDEKTTSQPNTHITSALTSMKAKGNFTNKTNNSTDLHIASTPTSQD
DATPSIPSVQTPNYNTNAPTRTLTSLHIEEGPSNSTTSEKATASTLSHNSHKNDTGGI
YTTLNKTTQLPSTNKPTNSQAKSSTKPRVETHNKTTSNPAISLTDSADVPQRPREPTL
PPIFRPPASKNRYLEKQLVIGLLTAVALTCGLITLFHYLFFR"
CDS 3194. .6592
/note="ORF 06, major ssDNA binding protein homolog"
/codon_start=1
/protein_id="AAB62603.1"
/db_xref="GI:2246478"
/translation="MALKGPQTLEENIGSAAPTGPCGYLYAYLTHNFPIGEASLLGNG
YPEAKVFSLPLLHGLTVESDFPLNVKAVHKKIDATTASVKLTSYHREAIVFHNTHLFQ
PIFQGKGLEKLCRESRELFGFSTFVEQQHKGTLWSPEACPQLPCANEIFMAVIVTEGF
KERLYGGKLVPVPSQTTPVHIGEHQAFKIPLYDEDLFGPSRAQELCRFYNPDISRYLH
DSIFTGIAQALRVKDVSTVIQASERQFVHDQYKIPKLVQAKDFPQCASRGTDGSTLMV
IDSLVAELGMSYGLSFIEGPQDSCEVLNYDTWPIFENCETPDARLRALEVWHAEQALH
IGAQLFAANSVLYLTRVAKLPQKNQRGDANMYNSFYLQHGLGYLSEATVKENGASAFK
GVPVSALDGSSYTLQHLAYASSFSPHLLARMCYYLQFLPHHKNTNSQSYNVVDYVGTA
APSQMCDLCQGQCPAVCINTLFYRMKDRFPPVLSNVKRDPYVITGTAGTYNDLEILGN
FATFRREREEEGNPVEDAPKYTYWQLCQNITEKLASMGISEGGDALRTLIVDIPSFVKV
FKGIDSTVEAELLKFINCMIKNNYNFRENIKSVHHILQFACNVYWQAPCPVFLTLYYK
SLLTVIQDICLTSCMMYEQDNPAVGIVPSEWLKMHFQTMWTNFKGACFDKGAITGGEL
KIVHQSMFCDLFDTDAAIGGMFAPARMQVRIARAMLMVPKTIKIKNRIIFSNSTGAES
IQAGFMKPASQRDSYIVGGPYMKFLNALHKTLFPSTKTSALYLWHKIGQTTKNPILPG
VSGEHLTELCNYVKASSQAFEEINVLDLVPDTLTSYAKIKLNSSILRACGQTQFYATT
LSCLSPVTQLVPAEEYPHVLGPVGLSSPDEYRAKVAGRSVTIVQSTLKQAVSTNGRLR
PIITVPLVVNKYTGSNGNTNVFHCANLGYFSGRGVDRNLRPESVPFKKNNVSSMLRKR
```
```

HVIMTPLVDRLVKRIVGINSGEFEAEAVKRSVQNVLEDRDNPNLPKTVVLELVKHLGS
SCASLTEEDVIYYLGPYAVLGDEVLSLLSTVGQAGVPWTAEGVASVIQDIIDDCELQF
VGPEEPCLIQGQSVVEELFPSPGVPSLTVGKKRKIASLLSDLDL"

CDS 6609. .8696
/note="ORF 7, processing and transport protein homolog (ICP18.5)"
/codon_start=1
/protein_id="AAB62620.1"
/db_xref="GI:2246495"
/translation="MAKELAAVYADVSALAMDLCLLSYADPATLDTKSLALTTGKFQS
LHGTLLPLLRRQNAHECSGLSLELEHLLENVADALATLGVCTSRKLSPEEHFSLLHLD
TTCNKHRSVRFNFYGNWALELKLSLINDVEIFFKRLSSVFYCIGSGSALEGLGEVLRF
VGKLRGISPVPGPDLYVSNLPCLECLQEVCLTPNQGTSLQAMLPDTACSHICTPACGE
PVRGLFENELKQLGLQTPESIPTTPCQSRVRQDDEIRQSSLMAVGDHHIFGEVTRSVL
EISNLIYWSSGHSDATCDGDRDCSHLASLFTHEADMHKRRVDLAGCLGERGTPKHFFD
CFRPDSLETLFCGGLFSSVEDTIESLQKDCSSAFYQQVNYTTALQKQNEFYVRLSKLL
AAGQLNLGKCSTESCPSEARRQLVGGGKPEEVLRDAKHRQELYLQKVARDGFKKLSDC
IRHQGHILSQTLGLRLWGSVIYNEASALQNHFLHRAQFISLPWQDLTVDCPTRFENSK
YIKNSLYCQRLGREHVEILTLEFYKLITGPLSKRHTLFPSPPNVTLAQCFEAAGMLPH
QKMMVSEMIWPSIEPKDWIEPNFNQFYSFENQDINHLQKRAWEYIRELVLSVSLYNRT
WERELKILLTPQGSPGFEEPKPAGLTTGLYLTFETSAPLVLVDKKYGWIFKDLYALLY
HHLQLSNHNDSQV"

CDS 8680. .11217
/note="ORF 08, glycoprotein B (gB) homolog"
/codon_start=1
/protein_id="AAB62592.1"
/db_xref="GI:2246467"
/translation="MTPRSRLATLGTVILLVCFCAGAAHSRGDTFQTSSSPTPPGSSS
KAPTKPGEEASGPKSVDFYQFRVCSASITGELFRFNLEQTCPDTKDKYHQEGILLVYK
KNIVPHIFKVRRYRKIATSVTVYRGLTESAITNKYELPRPVPLYEISHMDSTYQCFSS
MKVNVNGVENTFTDRDDVNTTVFLQPVEGLTDNIQRYFSQPVIYAEPGWFPGIYRVRT
TVNCEIVDMIARSAEPYNYFVTSLGDTVEVSPFCYNESSCSTTPSNKNGLSVQVVLNH
TVVTYSDRGTSPTPQNRIFVETGAYTLSWASESKTTAVCPLALWKTFPRSIQTTHEDS
FHFVANEITATFTAPLTPVANFTDTYSCLTSDINTTLNASKAKLASTHVPNGTVQYFH
TTGGLYLVWQPMSAINLTHAQGDSGNPTSSPPPSASPMTTSASRRKRRSASTAAAGGG
GSTDNLSYTQLQFAYDKLRDGINQVLEELSRAWCREQVRDNLMWYELSKINPTSVMTA
IYGRPVSAKFVGDAISVTECINVDQSSVNIHKSLRTNSKDVCYARPLVTFKFLNSSNL
FTGQLGARNEIILTNNQVETCKDTCEHYFITRNETLVYKDYAYLRTINTTDISTLNTF
IALNLSFIQNIDFKAIELYSSAEKRLASSVFDLETMFREYNYYTHRLAGLREDLDNTI
DMNKERFVRDLSEIVADLGGIGKTVVNVASSVVTLCGSLVTGFINFIKHPLGGMLMII
IVIAIILIIFMLSRRTNTIAQAPVKMIYPDVDRRAPPSGGAPTREEIKNILLGMHQLQ
QEERQKADDLKKSTPSVFQRTANGLRQRLRGYKPLTQSLDISPETGE"

CDS 11344. .14382
/note="ORF 09, DNA polymerase homolog"
/codon_start=1
/protein_id="AAB62593.1"
/db_xref="GI:2246468"
/translation="MDFFNPFIDPTRGGPRNTVRQPTPSQSPTVPSETRVCRLIPACF
QTPGRPGVVAVDTTFPPTYFQGPKRGEVFAGETGSIWKTRRGQARNAPMSHLIFHVYD
IVETTYTADRCEDVPFSFQTDIIPSGTVLKLLGRTLDGASVCVNVFRQRCYFYTLAPQ
GVNLTHVLQQALQAGFGRASCGFSTEPVRKKILRAYDTQQYAVQKITLSSSPMMRTLS
DRLTTCGCEVFESNVDAIRRFVLDHGFSTFGWYECSNPAPRTQARDSWTELEFDCSWE
DLKFIPERTEWPPYTILSFDIECMGEKGFPNATQDEDMIIQISCVLHTVGNDKPYTRM
LLGLGTCDPLPGVEVFEFPSEYDMLAAFLSMLRDYNVEFITGYNIANFDLPYIIARAT
QVYDFKLQDFTKIKTGSVFEVHQPRGGSDGGNFMRSQSKVKISGIVPIDMYQVCREKL
SLSDYKLDTVAKQCLGRQKDDISYKDIPPLFKSGPDGRAKVGNYCVIDSVLVMDLLLR
FQTHVEISEIAKLAKIPTRRVLTDGQQIRVFSCLLEAAATEGYILPVPKGDAVSGYQG
ATVISPSPGFYDDPVLVVDFASLYPSIIQAHNLCYSTLIPGDSLHLHPHLSPDDYETF
VLSGGPVHFVKKHKRESLLAKLLTVWLAKRKEIRKTLASCTDPALKTILDKQQLAIKV
TCNAVYGFTGVASGILPCLNIAETVTLQGRKMLERSQAFVEAISPERLAGLLRRPIDV
SPDARFKVIYGDTDSLFICCMGFNMDSVSDFAEELASITTNTLFRSPIKLEAEKIFKC
LLLLTKKRYVGVLSDDKVLMKGVDLIRKTACRFVQEKSSQVLDLILREPSVKAAAKLI
SGQATDWVYREGLPEGFVKIIQVLNASHRELCERSVPVDKLTFTTELSRPLADYKTQN
LPHLTVYQKLQARQEELPQIHDRIPYVFVDAPGSLRSELAEHPEYVKQHGLRVAVDLY
FDKLVHAVANIIQCLFQNNTSATVAMLYNFLDIPVTFPTP"

CDS 14500. .15756
/note="ORF 10, herpesvirus saimiri ORF 10 homolog, conserved in other gamma-herpesviruses"
/codon_start=1
/protein_id="AAB62594.1"
/db_xref="GI:2246469"
/translation="MQTEATFILGDWEITVSNCRFTCSSLTCGPLYRSSGDYTRLRIP

Query Match 19.1%; Score 191.4; DB 14; Length 133661;
Best Local Similarity 56.3%; Pred. No. 1.1e-21;
Matches 425; Conservative 0; Mismatches 316; Indels 14; Gaps 3;

```
Qy        68 CGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGA 127
             | || ||||| | |||    | ||| ||  | ||     ||||   |  ||| | || ||
Db     24992 CCCCTCCGGGGGAGGGTGGCGCGCCGGGATCCCCCTCCGGGGGAGGGTGGCGCGCCGGGA 24933

Qy       128 CCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGC----GAGGGGAGCGGGGGACAGG 183
              |      ||  |||||   || |   |||||   | |    | || | | || |  | ||
Db     24932 TCCCCCTCCGGGGGAGGGTGGCGCGCCGGGATCCCCCTCCGGGGGAGGGTGGCGCGCCGG 24873

Qy       184 GGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCC------TGCCTGCTCGCT 237
             | | | | | | | |   |||   | |||||  | |  ||        |  || |||
Db     24872 GATCCCCCTCCGGGGGAGGGTGGCGCGCCGGGATCCCCCTCCGGGGGAGGGTGGCGCGCC 24813

Qy       238 GGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGAC 297
             ||  | |       |||||  || |  |  | | || |   || |   |  | | |
Db     24812 GGGATCCCCCTCCGGGGGAGGGTGGCGCGCCGGGATCCCCCTCCGGGGGGGGGCGCGCCG 24753

Qy       298 AGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGG 357
               |   | | |||||  ||| |    ||  |||| |  |    |  |  ||||| |
Db     24752 GGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCG 24693

Qy       358 GGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCT 417
             || |    ||| |||| |  |       ||| |||  | || || || |  |||| |
Db     24692 GGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCG 24633

Qy       418 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 477
             || || || |  |||| | || || || |  |||| | || || || |  |||| |
Db     24632 GGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCG 24573

Qy       478 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 537
             || || || |  |||| | || || || |  |||| | || || || |  |||| |
Db     24572 GGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCG 24513

Qy       538 GGCCTGCT----TGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCC 593
             || || ||    |||||  ||| |    ||| |||| |  |       ||  |   |
Db     24512 GGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCG 24453

Qy       594 TGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 653
               | ||  |||  || |    ||||| ||||| ||||    ||||| ||||| ||||
Db     24452 GGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCG 24393

Qy       654 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTA 713
             |||||  |||||  ||||    |||||  |||||  ||||    ||||||||||| ||||
Db     24392 GGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCG 24333

Qy       714 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 773
             |||||  |||||  ||||    |||||||||||  ||||    |||||  |||||  ||||
Db     24332 GGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCG 24273

Qy       774 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 808
             |||||  |||||  ||||    |||||  |||||  |||
Db     24272 GGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGC 24238
```

RESULT 25
AC110455/c
LOCUS AC110455 183863 bp DNA linear HTG 15-NOV-2002
DEFINITION Rattus norvegicus clone CH230-276C17, WORKING DRAFT SEQUENCE, 3 unordered pieces.
ACCESSION AC110455
VERSION AC110455.6 GI:25006850
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 183863)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,

```
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 183863)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 183863)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 15, 2002 this sequence version replaced gi:22856759.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GRIZ
                Center clone name: CH230-276C17
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 169880 bases at least Q40
                Consensus quality: 170862 bases at least Q30
                Consensus quality: 171472 bases at least Q20
                Estimated insert size: 169890; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   181513: contig of 181513 bp in length
            *   181514   181613: gap of unknown length
            *   181614   182673: contig of 1060 bp in length
            *   182674   182773: gap of unknown length
            *   182774   183863: contig of 1090 bp in length.
FEATURES             Location/Qualifiers
     source          1. .183863
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-276C17"
     misc_feature    1. .1195
                     /note="wgs_contig"
     misc_feature    180129. .181513
                     /note="wgs_contig"
BASE COUNT    51113 a  33333 c  34320 g  52915 t  12182 others
ORIGIN

  Query Match             19.0%;  Score 190.2;  DB 2;  Length 183863;
  Best Local Similarity   58.6%;  Pred. No. 1.6e-21;
  Matches  330;  Conservative    0;  Mismatches  233;  Indels    0;  Gaps    0;

Qy        219 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 278
              | ||||||||||||  ||  |||||| |||    | | | |   |||| | |||  ||
Db     169622 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGTGTGCCTGCCTGCCTACCTGCCTGCCT 169563

Qy        279 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 338
              ||||||  ||    | ||| |   |||||| |||      | | |   |||||| |||
Db     169562 GCCTGCCAGCCTGTGTGACTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 169503

Qy        339 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 398
              ||  ||||||.|||    | | | |   | ||||  || |||  |   | ||||  ||
Db     169502 GCCTGCCTGCCTGCCTGTGTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCGT 169443

Qy        399 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 458
              |||||| ||| |||  ||  |||||| ||| |||  ||  |||||| ||| |||  ||
Db     169442 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 169383

Qy        459 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 518
              |  ||| ||| |||  ||  |||||| ||| |||  ||  |||||| ||| |||  ||
Db     169382 GTGTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 169323

Qy        519 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTA 578
              |||||| ||  |||  ||  |||||| |||      | | ||  | |||| ||| |||
Db     169322 GCCTGCCTGTGTGCCTGCCTGCCTGCCTGCCAGCCTGTGTGCCTGCCTGCCTGCCT 169263
```

```
Qy        579 AGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 638
              | |   |||||||||   || ||   |   | ||||    ||||  |   | ||
Db     169262 GCCTACCTGCCTGCCTGCCTGCCTGCCTGTGTGCCTGCCTGCCTACCTGCCTGCCTGCCT 169203

Qy        639 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 698
              ||  | ||||  |   | ||||  | ||||  ||  | ||||  |.||||  |   | ||
Db     169202 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTGTGCCTGCCTGCCTGCCTGCCTGCCT 169143

Qy        699 CCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCT 758
              |  | || |  ||  | |||   | ||||  |   | || |  | ||||  |   | ||
Db     169142 GCCTGCCTCCCTGTGTGCCTGTGTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 169083

Qy        759 GCTGGGCTGCTAGTAGGGCTGCT 781
              ||  | ||||  |  |   ||||
Db     169082 GCCTGCCTGCCTGCTGCCATGCT 169060
```

RESULT 26
AC094809
LOCUS AC094809 234939 bp DNA linear HTG 09-MAY-2003
DEFINITION Rattus norvegicus clone CH230-4C12, *** SEQUENCING IN PROGRESS ***, 2 unordered pieces.
ACCESSION AC094809
VERSION AC094809.6 GI:30466564
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 234939)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 234939)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 234939)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 9, 2003 this sequence version replaced gi:24941692.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
-------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
Center project name: GBKX
Center clone name: CH230-4C12
-------------- Summary Statistics
Assembly program: Atlas;
Consensus quality: 216649 bases at least Q40
Consensus quality: 220061 bases at least Q30
Consensus quality: 222335 bases at least Q20
Estimated insert size: 232189; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 231801: contig of 231801 bp in length
* 231802 231901: gap of unknown length
* 231902 234939: contig of 3038 bp in length.
FEATURES Location/Qualifiers
source 1. .234939
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-4C12"
misc_feature 1. .1673
/note="wgs_contig"
misc_feature 226281. .228013
/note="wgs_contig"

```
     misc_feature    230341. .231801
                     /note="wgs_contig"
BASE COUNT    71368 a  44035 c  42626 g  65688 t  11222 others
ORIGIN

  Query Match           19.0%;  Score 189.8;  DB 2;  Length 234939;
  Best Local Similarity   57.5%;  Pred. No. 1.7e-21;
  Matches  341;  Conservative    0;  Mismatches  252;  Indels    0;  Gaps    0;

Qy        218 CGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 277
              | ||  | |||||||  |   ||||||  |||      | | |   ||||||  |||  |
Db     172236 CAGCGCCATGCCTGCCTGTGTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTG 172295

Qy        278 GGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 337
              ||||||  |||  |   | | |   ||||||  ||    | |||  |   ||||||  |||
Db     172296 TGCCTGCCTGCCTACCTGCCTGCCTGCCTGCCAGCCTGTGTGACTGCCTGCCTGCCTGCC 172355

Qy        338 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCT 397
              ||  |||||| |||      | | |   | ||||  || ||   |   | ||||  ||
Db     172356 TGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTGTGCCTGCCTGCCTGCC 172415

Qy        398 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 457
              ||||||  |||  |||   ||  ||  |||  |||  |||   ||  ||||||  |||  |||  ||
Db     172416 TGCCTGCCTGCCTGCCTGCCTGCGTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCC 172475

Qy        458 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 517
              ||||||  |||  |||   ||  |||||  |||  |||   ||  ||||||  |||  |||  ||
Db     172476 TGCCTGCCTGCCTGCCTGCCTGCCTGTGTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCC 172535

Qy        518 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCT 577
              ||||||  |||  |||   ||  ||||||  |||     | | | |    | ||||  |||  |||
Db     172536 TGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTGTGCCTGCCTGCCTGCCTGCCTGCC 172595

Qy        578 AAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGC 637
              |    |   |  |||||||||   ||  |     |   | ||||  | ||||  ||   | |
Db     172596 AGCCTGTGTGCCTGCCTGCCTGCCTGCCTACCTGCCTGCCTGCCTGCCTGCCTGTGTGCC 172655

Qy        638 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 697
              |||  | ||  |   |   | ||||  | ||||  |   | ||||  | ||||  |   | |
Db     172656 TGCCTGCCTACCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCC 172715

Qy        698 TCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGC 757
              |    | ||||  |   | ||||  | ||||  |     ||    | |||   |   | |
Db     172716 TGTGTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCTGTGTGCCTGTGTGCCTGCC 172775

Qy        758 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 810
              |||  | ||||  |   | ||||  | ||||  |   | ||||||   ||||
Db     172776 TGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCTGCCATGCTA 172828


RESULT 27
AC096116
LOCUS       AC096116               224055 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-10O16, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC096116
VERSION     AC096116.7  GI:30522727
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 224055)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 224055)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 224055)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24817844.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
```

```
            Center project name: GELV
            Center clone name: CH230-10O16
          ------------- Summary Statistics
            Assembly program: Atlas 3.0;
            Consensus quality: 196024 bases at least Q40
            Consensus quality: 199448 bases at least Q30
            Consensus quality: 201887 bases at least Q20
            Estimated insert size: 204330; sum-of-contigs estimation
            Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
          -------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 2 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1   222891: contig of 222891 bp in length
          *  222892   222991: gap of unknown length
          *  222992   224055: contig of 1064 bp in length.
FEATURES             Location/Qualifiers
     source          1. .224055
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-10O16"
     misc_feature    1. .1326
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    2494. .3408
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH308507"
     misc_feature    53120. .54604
                     /note="wgs_contig"
     misc_feature    123391. .126252
                     /note="wgs_contig"
     misc_feature    221867. .222775
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH308510"
BASE COUNT    54935 a  43967 c  46140 g  58898 t  20115 others
ORIGIN

  Query Match             18.9%;  Score 189.2;  DB 2;  Length 224055;
  Best Local Similarity   55.0%;  Pred. No. 2.2e-21;
  Matches  371;  Conservative    0;  Mismatches  303;  Indels    0;  Gaps    0;

Qy        214 GGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 273
              ||| |   ||  ||  ||| |||||| ||| ||||   ||    | ||   | | | |
Db     154481 GGTTGGTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGTTGATTGGTTGGTTGGTT 154540

Qy        274 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCT 333
              | ||||  ||| ||||    |   || |||  ||| ||||   ||   || |||  |||
Db     154541 GGTTGGCTGGCTGGCTGGTTAGTTGGTTGGCTGGCTGGCTGGTTGGTTGGCTGGCT 154600

Qy        334 TGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCT 393
              ||| |||||   | | | ||    |   || ||   ||||||| | |||| ||   |||
Db     154601 GGCTGGCTGGTTGGTTGGTTGGTTAGTTGGTTGGCTGGCTGGTTAGTTGGTTGGCT 154660

Qy        394 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 453
              ||||||  ||| ||| ||| | ||    | | | | | | | ||||  ||| ||| | |
Db     154661 GGCTGGCTGGCTGGCTGGCTGGTTGATTGGTTGGTTGGTTGGCTGGCTAGCTGGTT 154720

Qy        454 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 513
              | |||   ||| ||| |   |||||| | | | | | | |||   ||| ||| | |
Db     154721 AGTTGGTTGGCTGGCTGGGCGGCTGGCTGGTTGGTTGGTTAGTTGGTTGGCTGGCTGGGT 154780

Qy        514 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCT 573
              ||||||  ||| ||| ||| | |||   | | | ||   ||   |  ||| ||| |||
Db     154781 GGCTGGCTGGCTGGCTGGCTGGTTGGTTGGTTGGTTGGTTGGCTGGGTGGCTGGCT 154840

Qy        574 TGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTA 633
              | | |  ||  |||  |   ||  ||||||| | | || ||   | |||   ||| |
Db     154841 GGTTGATTGGTTGGTTGGTTGGCTGGATGGCTGGTTGGTTGGTTGGTTGGCTGGAT 154900

Qy        634 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 693
              ||   | | |   | | |  ||   | |||   ||| |  ||   |||||   | | ||
Db     154901 GGCTGGTTAGTTGGTTGGCTGGCTGGGTGGCTGGCTGGCTGGCTGGCTGGTTGGTTGGTT 154960

Qy        694 GGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTA 753
              ||     |||   | | |  ||   |||||   | | || ||    |||||  ||| |
Db     154961 GGTTGGTTGGTTGGTTGGCTGGCTAGCTGGTTAGTTGGTTGGCTGGCTGGGTGGCTGGCT 155020

Qy        754 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 813
              ||   | |||   | | || ||   ||||||  ||| |  ||   |||||   ||| ||
Db     155021 GGTTGGTTGGTTAGTTGGTTGGCTGGCTGGGTGGCTGGCTGGCTGGCTGGCTGGTT 155080

Qy        814 GGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGC 873
              ||   |   | ||| | | | ||| | | ||| |  ||     | |   | |  | ||
Db     155081 GGTTGGTTGGTTGGTTGGCTGGGTGGCTGGCTGGTTGATTGGTTGGTTGGTTGGCT 155140

Qy        874 TGCTGTGCAGCTGG 887
              | ||    | |||
Db     155141 GGATGGCTGGTTGG 155154


RESULT 28
AC114591/c
LOCUS       AC114591              208740 bp    DNA     linear   HTG 24-FEB-2003
DEFINITION  Mus musculus clone RP23-478G2, WORKING DRAFT SEQUENCE, 7 ordered
            pieces.
ACCESSION   AC114591
VERSION     AC114591.5  GI:28475804
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 208740)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-478G2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 208740)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 208740)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
```

```
          Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
          Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
          Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
          Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
          Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
          Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
          Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
          Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
          Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
          Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
          Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
          Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
          O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
          Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
          Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
          Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
          Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
          Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
          Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE     Direct Submission
JOURNAL   Submitted (24-FEB-2003) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Feb 24, 2003 this sequence version replaced gi:21392539.
          All repeats were identified using RepeatMasker:
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          -------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
          -------------- Project Information
              Center project name: L24156
              Center clone name: 478_G_2
          -------------- Summary Statistics
              Sequencing vector: Plasmid; n/a; 100% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.960731
              Consensus quality: 207158 bases at least Q40
              Consensus quality: 207792 bases at least Q30
              Consensus quality: 208011 bases at least Q20
              Insert size: 210000; agarose-fp
              Insert size: 208140; sum-of-contigs
              Quality coverage: 11.2 in Q20 bases; agarose-fp
              Quality coverage: 11.3 in Q20 bases; sum-of-contigs
          --------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 7 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *        1    11975: contig of 11975 bp in length
          *    11976    12075: gap of 100 bp
          *    12076    14622: contig of 2547 bp in length
          *    14623    14722: gap of 100 bp
          *    14723   125098: contig of 110376 bp in length
          *   125099   125198: gap of 100 bp
          *   125199   139380: contig of 14182 bp in length
          *   139381   139480: gap of 100 bp
          *   139481   158709: contig of 19229 bp in length
          *   158710   158809: gap of 100 bp
          *   158810   195661: contig of 36852 bp in length
          *   195662   195761: gap of 100 bp
          *   195762   208740: contig of 12979 bp in length.
FEATURES             Location/Qualifiers
     source          1. .208740
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-478G2"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .11975
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    12076. .14622
                     /note="assembly_fragment"
     misc_feature    14723. .125098
                     /note="assembly_fragment"
     misc_feature    125199. .139380
                     /note="assembly_fragment"
     misc_feature    139481. .158709
                     /note="assembly_fragment"
     misc_feature    158810. .195661
                     /note="assembly_fragment"
     misc_feature    195762. .208740
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    59319 a  44577 c  46204 g  58040 t    600 others
ORIGIN

  Query Match             18.7%;  Score 186.8;  DB 2;  Length 208740;
  Best Local Similarity   56.3%;  Pred. No. 5.3e-21;
  Matches  350;  Conservative    0;  Mismatches  272;  Indels    0;  Gaps    0;

Qy        230 TGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCT 289
              |||| | | ||  |  | ||| |  |   |   ||  | ||||| | | ||  | |||| 
Db     148302 TGCTGGGTTGCTGGGATACTGGGTTGCTGGGTTGCTGGGTTGCTGGGTTGCTGGGTTGCT 148243

Qy        290 GAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCT 349
              | |  |  ||  ||  |  | ||| | ||   |  |||  | ||||| |    |  | |
Db     148242 GGGTTGCTAGGTTGCTGGGATACTGGGTGGCTGGGTGGCTGGGTTGCTGGGATACTGGGT 148183

Qy        350 TGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCT 409
              ||||| | ||   |  ||   | ||||| | | |  |||  ||| | ||||  | |||||
Db     148182 TGCTGGGTGGCTGGGTGGCTGGGTGGCTGGGTGGCTGGGTGGCTGGATGGCTGGGTTGCT 148123

Qy        410 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 469
              | | |||||  |||| | || || | | ||  | ||||| | | |||||  | || | |
Db     148122 GGGTTGCTGGGTTGCTGGGTTACTGGGTTGCTGGGTTGCTAGGTTGCTGGGATACTGGGT 148063

Qy        470 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 529
              |||  | | ||  | |||||  | |  |||||  | || | |||||  | | ||  | |||||
Db     148062 GGCTGGGTTGCTGGGTTGCTGGTTTGCTGGGTTACTGGGTTGCTGGGTTGCTGGGTTGCT 148003

Qy        530 TGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTA 589
              | | |||||  | || | ||   ||   |  ||| |||| | || ||  |  |  ||
Db     148002 AGGTTGCTGGGATACTGGGTGGCTGGGTTGCTGGGTTGCTGGGTTACTGGGTTGCTGGGT 147943

Qy        590 CGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 649
              ||  |  |  ||| ||| | |   |  | | |   ||| | |||| | || |  |  |||
Db     147942 TGCTGGGTTACTGGGTTGCTGGGTTGCTGGGTTGCTAGGTTGCTGGGATACTGGGTGGCT 147883

Qy        650 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCT 709
              .  | ||  | |||  | | | ||| | ||||| |||| |  ||     |||  | |
Db     147882 GGGTGGCTGGGTGGCTGGGTTGCTGGGATACTGGGTTGCTGGGTGGCTGGGTGGCTGGGT 147823

Qy        710 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 769
              |  ||  |||||| ||| |  ||     |||  | |||  ||| |||||||| | ||
Db     147822 GGCTGGATGGCTGGGTGGCTGGGTGGCTGGGTGGCTGGGTAGCTGGGTTGCTGGGATACT 147763

Qy        770 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCT 829
              |  |  | |||  | | |  ||| |||||  ||| |  ||   |  |||| |
Db     147762 GGGTTGCTGGGTGGCTGGGTGGCTGGGTGGCTGGGTGGCTGGGTGGCTGGGTGGCTGGGT 147703

Qy        830 TGCTTGCTTGCTTGCTAGTGGG 851
              ||| |  |||| | | |  ||
Db     147702 GGCTGGGATGCTGGGTTGCTGG 147681


RESULT 29
```

```
AL844846
LOCUS       AL844846              209013 bp    DNA     linear   ROD 05-MAR-2003
DEFINITION  Mouse DNA sequence from clone RP23-358I19 on chromosome 2, complete
            sequence.
ACCESSION   AL844846
VERSION     AL844846.11  GI:27902000
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 209013)
  AUTHORS   Bates,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 25, 2003 this sequence version replaced gi:25955774.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to confirm this sequence.  Sequence data
            from the whole genome shotgun alone has only been used where it has
            a phred quality of at least 30.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-358I19 is
            from the RPCI-23 Mouse BAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .209013
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-358I19"
                     /clone_lib="RPCI-23"
BASE COUNT    59522 a  49671 c  46169 g  53651 t
ORIGIN

  Query Match           18.5%;  Score 185.4;  DB 10;  Length 209013;
  Best Local Similarity  63.2%;  Pred. No. 8.8e-21;
  Matches  285;  Conservative    0;  Mismatches  166;  Indels    0;  Gaps    0;

Qy        402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461
              |||||||||| | ||| || ||| |||||||| ||| || ||||||||||| ||| ||
Db     152500 TGCTTGCTTGTTTGCTTGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 152559

Qy        462 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 521
              |||||||||||| | | || |||||||||||| ||| || |||||||||||| ||| |
Db     152560 TGCTTGCTTGCTTGTTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGTT 152619

Qy        522 TGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGG 581
              |||||||||||| | | || || ||| |      |   ||  |  ||||||||||||
Db     152620 TGCTTGCTTGCTTGTTTGCTTGTTTGTTTGCTTGTTTGTTTGCTTGCTTGCTTGCTTGTT 152679

Qy        582 GGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCT 641
              |   |   || ||| || | ||||| | |    |  |||| |  || | ||  |  ||||
Db     152680 TGCTTGCTTGCTTGCTTGTTTGCTTGCTTGCTTGTTTGCTTGCTTGTTTGTTTGTTTGCT 152739

Qy        642 GGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCT 701
              |  |||| ||  |  |||| |  |||| ||  |  |||| |  |||| |   |  | |
Db     152740 TGCTTGCTTGTTTGCTTGCTTGCTTGCTTGTTTGCTTGCTTGCTTGCTTGCTTGTTTGTT 152799

Qy        702 GGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCT 761
              |  ||||  |  |  || | |  || | |   |  | || |  || | ||  |  ||||
Db     152800 TGCTTGCTTATTTGTTTGTTTGCTTGTTTGCTTGCTTGCTTGTTTGTTTGTTTGCTTGCT 152859

Qy        762 GGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCC 821
              |  |||| |   |  || | |  |||| |   |  || | |  |||| ||  |  |||
Db     152860 TGTTTGCTTGCTTGCTTGTTTGCTTGCTTGCTTGCTTGTTTGCTTGCTTGTTTGTTTGCT 152919

Qy        822 TGCTGGCTTGCTTGCTTGCTTGCTAGTGGGG 852
              |||| | ||| ||| |||  || || ||| |
Db     152920 TGCTTGTTTGTTTGTTTGTCTGTTACTGGTG 152950


RESULT 30
AC094484/c
LOCUS       AC094484              235855 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-4B15, WORKING DRAFT SEQUENCE.
ACCESSION   AC094484
VERSION     AC094484.6  GI:30466820
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 235855)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
```

```
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 235855)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 235855)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 9, 2003 this sequence version replaced gi:24818868.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GASZ
                Center clone name: CH230-4B15
            -------------- Summary Statistics
                Assembly program: Atlas;
                Consensus quality: 224879 bases at least Q40
                Consensus quality: 227006 bases at least Q30
                Consensus quality: 228555 bases at least Q20
                Estimated insert size: 237610; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   235855: contig of 235855 bp in length.
FEATURES            Location/Qualifiers
     source         1. .235855
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-4B15"
     misc_feature   2767. .3585
                    /note="clone_boundary
                    clone_end:T7
                    site:EcoRI
                    end_sequence:BH303847"
     misc_feature   complement(231726. .232568)
                    /note="clone_boundary
                    clone_end:Sp6
                    site:EcoRI
                    end_sequence:BH303848"
BASE COUNT    61996 a  50589 c  50777 g  66609 t   5884 others
ORIGIN

  Query Match            18.3%;  Score 183.4;  DB 2;  Length 235855;
  Best Local Similarity  56.2%;  Pred. No. 1.8e-20;
  Matches  385;  Conservative    0;  Mismatches  296;  Indels    4;  Gaps    2;

Qy        194 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 253
              ||| ||| |||  | |      ||  | ||||||||||||  ||  |||||| |||
Db     186287 TGCCTGCCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGTC 186228

Qy        254 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG---GACAGTAGGCCTGCTT 310
              | | |   ||||   |||  ||  | |||| ||||||  |   | |  |  |||||  |
Db     186227 TGCCTGCCTGCCTTTCTGCCTGCCTGTCTGCCTGCTGACTGCCTGCCTTTCTGCCTGTCT 186168

Qy        311 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 370
              |       | | |   |||||| ||   ||  | |||  |||       | | |   | ||
Db     186167 GTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGGCTGCCTGCCTGCCTGCCTGTCT 186108

Qy        371 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 430
              |   ||  |||   |  | |||||  |||||  ||||  || |||  | |||  ||| |||
Db     186107 GTCTGCCTGCCTTTCTGCCTGCTGACTGGCTGGCTTTCTGGCTGTCCGGCTGGCTGGCTG 186048

Qy        431 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              |||  ||||||   ||  |||  |||  ||||||  |||  |||  ||  |||||  |||  ||
Db     186047 GCTGGCTGGCTGTCTGGCTGGCTGGCTGGCTGGCTGCCTGCCTGGCTGGCTGACTG 185988

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              ||  |||  |  |||  |||  |||  |||  |||||  |||  ||   ||  |||||  |||
Db     185987 CCTGCCTGCCTGGCTGGCTGGCTGGCT-GCCTGGCTGCCTGGCTGCCTGCCTGGCTGCCT 185929

Qy        551 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 610
              |  |  |    |||||  ||  |||      |||  |   |||||||||   |  ||
Db     185928 GCCTGCCTGCCTTGCTGCCTGGCTGCCTGCCTGACTGCCTGCCTGCCTGCCTGTCTGTCT 185869

Qy        611 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 670
              |   | ||||  | ||||  |     ||||  | |||   |   | ||||  | ||||
Db     185868 GCCTGCCTGCCTGCCTGCCTGCCTTTCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGCCT 185809

Qy        671 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 730
              ||  | ||||  | ||||  ||  | || |  | ||||   |  | |||   | ||||
Db     185808 GTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTTTCTGCCTGTCTGTCTGCCT 185749

Qy        731 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 790
              | || |  | ||||   |  | ||||  | ||||   |   | ||||  | ||||
Db     185748 ACCTGCCTGCCTGTCTGCCTTTCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCT 185689

Qy        791 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGG 850
              |   | ||||  | ||||  |   | ||| ||||   | ||| ||| ||| ||  |
Db     185688 GCCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTGCCTACCTGCCTGCCTGCCTGTCTGCCT 185629

Qy        851 GGCCGCTTGCCTGCTACTAGGGCTG 875
              | | |  |||||||   |  | |||
Db     185628 GCCTGACTGCCTGCCTTTCTGCCTG 185604


RESULT 31
AC097040
LOCUS       AC097040              230500 bp    DNA     linear   HTG 13-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-66A9, WORKING DRAFT SEQUENCE.
```

```
ACCESSION   AC097040
VERSION     AC097040.7  GI:24942620
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 230500)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 230500)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 230500)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 13, 2002 this sequence version replaced gi:23664821.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GHYK
                Center clone name: CH230-66A9
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 218247 bases at least Q40
                Consensus quality: 220700 bases at least Q30
                Consensus quality: 222115 bases at least Q20
                Estimated insert size: 223843; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   230500: contig of 230500 bp in length.
FEATURES             Location/Qualifiers
     source          1. .230500
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-66A9"
     misc_feature    1. .1087
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    2577. .3456
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BH264043"
BASE COUNT    57052 a  55956 c  55370 g  54914 t   7208 others
ORIGIN

  Query Match             18.0%;  Score 180.4;  DB 2;  Length 230500;
  Best Local Similarity   53.9%;  Pred. No. 5.3e-20;
  Matches  370;  Conservative    0;  Mismatches  316;  Indels    0;  Gaps    0;

Qy        191 GTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTG 250
              || ||  ||  |||  | |      |   | ||||||| ||||  |   |||||| |||
Db     101810 GTCTGTCTGTCTGTCTGTCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCT 101869

Qy        251 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 310
                | | ||  |||||  |||  |   |||||  |||      | | |   | ||   |
Db     101870 GCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGCCTGTCTGCCTGTCTCTCT 101929

Qy        311 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 370
              ||      | | ||  |||||  |||  |   |||||  |||      | | ||  | ||
Db     101930 GCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCT 101989
```

```
Qy        371 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 430
              ||  |  ||   ||  | |||   ||  |||||| ||  |||  |   | |||| ||| |
Db     101990 GCCTGTCTGTCTGTCTGCCTGTCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTGCCT 102049

Qy        431 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              ||  |   ||||  ||| ||   ||  | |||| ||  ||   |   ||||| ||  |
Db     102050 GCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGTCT 102109

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              |   |   | |||  ||| ||   ||  | |||| ||| |||  ||  |||||  |||
Db     102110 GTCTGTCTGTCTGTCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCT 102169

Qy        551 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 610
                | |  ||  | |||  ||| |||       | | ||  ||||| |||   |  ||
Db     102170 GTCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGTCTGTCT 102229

Qy        611 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 670
              ||  | |||   | ||||  |   | ||||  | ||||  ||  | ||||  | ||||
Db     102230 GTCTGCCTGTCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGTCTGCCTGCCTGCCT 102289

Qy        671 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 730
              ||  | ||||  | ||||  |   | ||  |  | ||||  ||  | |||   | |||
Db     102290 GTCTGCCTGCCTGTCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGTCTGTCTGTCTGTCT 102349

Qy        731 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 790
              |   | ||  |  | ||||  ||  | |||   | |||   ||  | ||||  | |||
Db     102350 GCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGTCTGCCTGCCTGCCTGTCT 102409

Qy        791 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGG 850
              |   | ||||  | ||||  ||  | ||||||||  |  ||| ||  ||| ||   |
Db     102410 GCCTGTCTGCCTGTCTGCCCGTCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCT 102469

Qy        851 GGCCGCTTGCCTGCTACTAGGGCTGC 876
              | | || ||||||   |  | ||||
Db     102470 GTCTGCCTGCCTGTCTGTCTGCCTGC 102495
```

RESULT 32
AC142075/c
LOCUS AC142075 258132 bp DNA linear HTG 13-MAY-2003
DEFINITION Rattus norvegicus clone CH230-140G17, WORKING DRAFT SEQUENCE, 5 unordered pieces.
ACCESSION AC142075
VERSION AC142075.3 GI:30579162
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 258132)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 258132)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 258132)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 13, 2003 this sequence version replaced gi:29469497.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
-------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
Center project name: GJTN
Center clone name: CH230-140G17
-------------- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 228398 bases at least Q40
Consensus quality: 231939 bases at least Q30
Consensus quality: 234817 bases at least Q20
Estimated insert size: 242804; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently

```
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    10188: contig of 10188 bp in length
            *    10189    10288: gap of unknown length
            *    10289   252038: contig of 241750 bp in length
            *   252039   252138: gap of unknown length
            *   252139   253967: contig of 1829 bp in length
            *   253968   254067: gap of unknown length
            *   254068   255789: contig of 1722 bp in length
            *   255790   255889: gap of unknown length
            *   255890   258132: contig of 2243 bp in length.
FEATURES             Location/Qualifiers
     source          1. .258132
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-140G17"
     misc_feature    2429. .5186
                     /note="wgs_contig"
BASE COUNT    59022 a  58568 c  58799 g  60346 t  21397 others
ORIGIN

  Query Match           18.0%;  Score 180.4;  DB 2;  Length 258132;
  Best Local Similarity   53.9%;  Pred. No. 5.2e-20;
  Matches  370;  Conservative    0;  Mismatches  316;  Indels    0;  Gaps    0;

Qy        191 GTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTG 250
              || ||  ||  |||  | |     |   | ||||||| ||||  |   |||||| |||
Db     149903 GTCTGTCTGTCTGTCTGTCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCT 149844

Qy        251 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 310
              | | ||  |||||  |||  |   |||||  |||       | | |   | ||   |
Db     149843 GCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGCCTGTCTGCCTGTCTCTCT 149784

Qy        311 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 370
              ||       | | ||  |||||  |||  |   |||||  |||       | | ||  | ||
Db     149783 GCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCT 149724

Qy        371 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 430
              ||  |  ||   ||  | |||   ||  |||||| ||  |||  |   | |||| ||| |
Db     149723 GCCTGTCTGTCTGTCTGCCTGTCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTGCCT 149664

Qy        431 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              ||  |   |||||  ||| ||   ||  | |||| ||  ||    |   ||||| ||  |
Db     149663 GCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGTCT 149604

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              |   |   | |||  ||| ||   ||  | |||| ||| |||  ||  |||||  |||
Db     149603 GTCTGTCTGTCTGTCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCT 149544

Qy        551 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 610
              | | ||  | |||  ||| |||       | | ||  |||||  |||   |  ||
Db     149543 GTCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGTCTGTCT 149484

Qy        611 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 670
              ||  | |||   | ||||  |   | ||||  | ||||  ||  | ||||  | ||||
Db     149483 GTCTGCCTGTCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGTCTGCCTGCCTGCCT 149424

Qy        671 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 730
              ||  | ||||  | ||||  |   | ||  |  | ||||  ||  | |||   | |||
Db     149423 GTCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGTCTGTCTGTCT 149364

Qy        731 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 790
              |   | ||  |  | |||  ||  | |||   | |||   ||  | ||||  | |||
Db     149363 GCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGTCTGCCTGCCTGCCTGTCT 149304

Qy        791 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGG 850
              |   | ||||  | ||||  ||  | ||||||||  |  ||| ||  ||| ||   |
```

```
Db     149303 GCCTGTCTGCCTGTCTGCCCGTCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCT 149244

Qy        851 GGCCGCTTGCCTGCTACTAGGGCTGC 876
              | | ||  ||||||    |  | ||||
Db     149243 GTCTGCCTGCCTGTCTGTCTGCCTGC 149218


RESULT 33
AC109525_0
WPCOMMENT
  Sequence split into 5 fragments  LOCUS AC109525 Accession AC109525
     Fragment Name      Begin     End
     AC109525_0             1     110000
     AC109525_1        100001     210000
     AC109525_2        200001     310000
     AC109525_3        300001     410000
     AC109525_4        400001     478059
LOCUS       AC109525              478059 bp    DNA     linear   HTG 11-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-255D23, *** SEQUENCING IN PROGRESS
            ***, 5 unordered pieces.
ACCESSION   AC109525
VERSION     AC109525.6  GI:23603814
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 478059)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
```

```
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 478059)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 478059)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 9, 2002 this sequence version replaced gi:22795167.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GPHZ
                Center clone name: CH230-255D23
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 380245 bases at least Q40
                Consensus quality: 385911 bases at least Q30
                Consensus quality: 389757 bases at least Q20
                Estimated insert size: 390792; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   235023: contig of 235023 bp in length
            *   235024   235123: gap of unknown length
            *   235124   246065: contig of 10942 bp in length
            *   246066   246165: gap of unknown length
            *   246166   474601: contig of 228436 bp in length
            *   474602   474701: gap of unknown length
            *   474702   475761: contig of 1060 bp in length
            *   475762   475861: gap of unknown length
            *   475862   478059: contig of 2198 bp in length.
FEATURES             Location/Qualifiers
     source          1. .478059
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-255D23"
     misc_feature    1. .2769
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    complement(9835. .10732)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:MboI
                     end_sequence:RXACK24TVB"
     misc_feature    235124. .236699
                     /note="wgs_contig"
     misc_feature    287702. .289205
                     /note="wgs_contig"
     misc_feature    350281. .351366
                     /note="wgs_contig"
     misc_feature    358886. .361234
                     /note="wgs_contig"
     misc_feature    406035. .407045
                     /note="wgs_contig"
BASE COUNT   110790 a  84236 c  86247 g 111645 t  85141 others
ORIGIN

  Query Match             18.0%;  Score 179.8;  DB 2;  Length 110000;
  Best Local Similarity   57.0%;  Pred. No. 8e-20;
  Matches  348;  Conservative    0;  Mismatches  262;  Indels    1;  Gaps    1;

Qy        273 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGC 332
              | ||||||    | ||||   ||   |  |||  ||| ||||   ||   |  |||  ||
Db      52847 TGGCTGGCTGATTGGCTGGTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGC 52906

Qy        333 TTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGC 392
              | ||| |||||   | | | ||   ||   |  ||   | ||| | | | || ||   |
Db      52907 TGGCTGGCTGGTTGGTTGGTTGGTTGGTTCGCTGGTTGGTTGGTTGGTTGGTTGGTTGGT 52966

Qy        393 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 452
              | | |||   | |  || ||| ||||||  ||| ||| ||| ||||||  ||| ||| ||
Db      52967 TGGTTGGTTGGTTGACTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGC 53026

Qy        453 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 512
              | |||||   | |||||| | | | ||||  ||| ||| ||| | |||   | | ||| ||
Db      53027 TGGCTGGTTGGATTGCTGGTTGGTTGGCTGGCTGGCTGGCTGGTTGGTTGGTTGGCTGGC 53086

Qy        513 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGC 572
              | ||||||  | | | | | | | |||   | | ||||   ||   || ||   ||| ||
Db      53087 TGGCTGGCTGGTTGGATGGTTGGTTGGTTGGCTGGCTGGCTGGTTGGCTGGCTGGC 53146

Qy        573 TTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGT 632
              | |||  |  | | |   |  ||  || | | ||| |      || || | || || | |
Db      53147 TGGCT-GGCTGGCTGGCTGGTTGGTTGGTTGGTTCGCTGGTTGGTTGGTTGGTTGG 53205

Qy        633 AGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGT 692
              || || | || || | |   || || | || ||   |  || || | || || | |
Db      53206 TTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGG 53265

Qy        693 AGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGT 752
              || |  |  |||| | |   || || | || || | |   || |  | || ||   |
Db      53266 TTGGTTGGTTCGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGACTGG 53325

Qy        753 AGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGT 812
              |||||   |||||  |   |||||  ||||| | | || || | || || | |
Db      53326 CTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGTTGGATGGTTGGTTGGTTGGTTGG 53385

Qy        813 AGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGG 872
              ||||| ||| | |  || |||  || |||   |   ||| | |||  || |  |||||
Db      53386 CTGGCTGGCTGGTTGGCTGGTTGGCTGGTTGGCTGGTTGGCTGGTTGGTTGATGATAGGG 53445

Qy        873 CTGCTGTGCAG 883
              ||   ||| ||
Db      53446 CTCAAGTGTAG 53456


RESULT 34
AC094443/c
```

```
LOCUS       AC094443              237019 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-4F6, *** SEQUENCING IN PROGRESS ***,
            3 unordered pieces.
ACCESSION   AC094443
VERSION     AC094443.6  GI:30466976
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 237019)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 237019)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 237019)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 9, 2003 this sequence version replaced gi:22772430.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GAPQ
                Center clone name: CH230-4F6
            -------------- Summary Statistics
                Assembly program: Atlas;
                Consensus quality: 229894 bases at least Q40
                Consensus quality: 231178 bases at least Q30
                Consensus quality: 232607 bases at least Q20
                Estimated insert size: 244941; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   231233: contig of 231233 bp in length
            *   231234   231333: gap of unknown length
            *   231334   233041: contig of 1708 bp in length
            *   233042   233141: gap of unknown length
            *   233142   237019: contig of 3878 bp in length.
FEATURES             Location/Qualifiers
     source          1. .237019
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-4F6"
     misc_feature    227374. .228903
                     /note="wgs_contig"
     misc_feature    228959. .231233
                     /note="wgs_contig"
BASE COUNT    73264 a  46066 c  45849 g  68157 t   3683 others
ORIGIN

  Query Match             18.0%;  Score 179.8;  DB 2;  Length 237019;
  Best Local Similarity   57.0%;  Pred. No. 6.6e-20;
  Matches  348;  Conservative    0;  Mismatches  262;  Indels    1;  Gaps    1;

Qy        273 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGC 332
              | ||||||    | ||||   ||   |  |||  ||| ||||   ||   |  |||  ||
Db      23735 TGGCTGGCTGATTGGCTGGTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGC 23676

Qy        333 TTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGC 392
              | ||| |||||   | | | ||   ||   |  ||   | ||| | | | || ||   |
Db      23675 TGGCTGGCTGGTTGGTTGGTTGGTTCGCTGGTTGGTTGGTTGGTTGGTTGGTTGGT 23616

Qy        393 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 452
              | | |||   | |  || ||| ||||||  ||| ||| ||| ||||||  ||| ||| ||
```

```
Db      23615 TGGTTGGTTGGTTGACTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGC 23556

Qy        453 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 512

Db      23555 TGGCTGGTTGGATTGCTGGTTGGTTGGCTGGCTGGCTGGCTGGTTGGTTGGCTGGC 23496

Qy        513 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGC 572

Db      23495 TGGCTGGCTGGTTGGATGGTTGGTTGGTTGGTTGGCTGGCTGGCTGGTTGGCTGGCTGGC 23436

Qy        573 TTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGT 632

Db      23435 TGGCT-GGCTGGCTGGCTGGTTGGTTGGTTGGTTCGCTGGTTGGTTGGTTGGTTGG 23377

Qy        633 AGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGT 692

Db      23376 TTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGG 23317

Qy        693 AGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGT 752

Db      23316 TTGGTTGGTTCGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGACTGG 23257

Qy        753 AGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGT 812

Db      23256 CTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGTTGGATGGTTGGTTGGTTGGTTGG 23197

Qy        813 AGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGG 872

Db      23196 CTGGCTGGCTGGTTGGCTGGTTGGCTGGTTGGCTGGTTGGCTGGTTGGTTGATGATAGGG 23137

Qy        873 CTGCTGTGCAG 883

Db      23136 CTCAAGTGTAG 23126
```

```
RESULT 35
AL935202/c
LOCUS       AL935202              189096 bp    DNA     linear   VRT 01-MAY-2003
DEFINITION  Zebrafish DNA sequence from clone CH211-197E11, complete sequence.
ACCESSION   AL935202
VERSION     AL935202.7  GI:30314908
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 189096)
  AUTHORS   Leongamornlert,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 1, 2003 this sequence version replaced gi:28446340.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Repeat names
            beginning 'Dr' were identified by the Recon repeat discovery system
            (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr'
            were identified by Rick Waterman (Stephen Johnson lab, WashU). For
            further information see
            http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml
            CH211-197E11 is from a CHORI-211 BAC library
            VECTOR: pTARBAC2.1.
FEATURES             Location/Qualifiers
     source          1. .189096
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-197E11"
                     /clone_lib="CHORI-211"
BASE COUNT    58203 a  34096 c  34460 g  62337 t
ORIGIN
```

Query Match 17.9%; Score 178.8; DB 5; Length 189096;
Best Local Similarity 54.2%; Pred. No. 1e-19;
Matches 363; Conservative 0; Mismatches 307; Indels 0; Gaps 0;

```
Qy        219 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 278

Db      99704 GGTTGACTGCTTGCTTGGTTGTTTGGTTGATTGGTTGGTTGGTTAGTCGGTTGCTTGCTT 99645

Qy        279 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 338

Db      99644 GCTTGCTAGTTTGTTTGTTTGCTTGGTTGATTGCTTGCTTGGTTGGTTGTTTGGTTGCTT 99585

Qy        339 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 398

Db      99584 GTTTGCTTGGTTGCTCAGTTGATTGCTTGGTTGATAGCTTGGTTGGTTGGTTGGGTAGTT 99525

Qy        399 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 458

Db      99524 GCTTGCTTGTTTGCTTGGTTGGTTGGTTGGTTGTTTGGTTGATTGCTTGGTTCGTTGGTT 99465

Qy        459 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 518

Db      99464 GATTGGTTGGTTGTTTGGTTGATTTCTTGGTTGGTTGGTTGATTGCTTGGTTAGTTTATT 99405

Qy        519 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTA 578

Db      99404 GGTTGGTTAGTTTCTCGCTTGCTTGCTTGTTTGCTTGGTTGGTTGGTTGGTTAGTTGCTT 99345

Qy        579 AGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 638

Db      99344 GCTTGTTTGTTTGCTTGTTTAGTTGGTTGGTTGCATGTTTGCTTGGTTGATTGCTTGGTT 99285

Qy        639 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 698

Db      99284 ATTTGGTTGGTTGGCTGGTTGGTTAGTTGGATGGTTGGTCGGTTGCTTGGTTGATGGCTT 99225

Qy        699 CCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCT 758

Db      99224 GCTTGCGTGTTTGCTTGCTTTGTTGGTTGGTTGGTTGACTGCTTGGTCGGTTGGTTGATT 99165

Qy        759 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 818

Db      99164 ACTTGGTTGGTTAGTTGATTGGTTGGTTGGTTGTTTGGTTGATTGCTTGGTTGGTTGGTT 99105

Qy        819 GCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTG 878

Db      99104 GGTTGCTTGCTTGGTTGGTTGGCTGGTTGTTTGGTCGATTGCTTGGTTGGTTGGTTGCTT 99045

Qy        879 TGCAGCTGGG 888
```

```
Db      99044 GGTTGGTTGG 99035


RESULT 36
AC112928
LOCUS       AC112928               223156 bp    DNA     linear   HTG 18-MAR-2003
DEFINITION  Mus musculus clone RP24-275P22, WORKING DRAFT SEQUENCE, 8 unordered
            pieces.
ACCESSION   AC112928
VERSION     AC112928.4  GI:29029390
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 223156)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-275P22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 223156)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 223156)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 18, 2003 this sequence version replaced gi:28209605.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
               Center project name: L18718
               Center clone name: 275_P_22
            -------------- Summary Statistics
               Sequencing vector: Plasmid; n/a; 100% of reads
               Chemistry: Dye-terminator Big Dye; 100% of reads
               Assembly program: Phrap; version 0.960731
               Consensus quality: 221447 bases at least Q40
               Consensus quality: 222092 bases at least Q30
               Consensus quality: 222332 bases at least Q20
               Insert size: 194000; agarose-fp
               Insert size: 222456; sum-of-contigs
               Quality coverage: 11.2 in Q20 bases; agarose-fp
               Quality coverage: 9.8 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5792: contig of 5792 bp in length
            *     5793     5892: gap of 100 bp
            *     5893     6990: contig of 1098 bp in length
            *     6991     7090: gap of 100 bp
            *     7091    12953: contig of 5863 bp in length
            *    12954    13053: gap of 100 bp
            *    13054    34245: contig of 21192 bp in length
            *    34246    34345: gap of 100 bp
            *    34346    60533: contig of 26188 bp in length
            *    60534    60633: gap of 100 bp
            *    60634   100582: contig of 39949 bp in length
            *   100583   100682: gap of 100 bp
            *   100683   192104: contig of 91422 bp in length
            *   192105   192204: gap of 100 bp
            *   192205   223156: contig of 30952 bp in length.
FEATURES             Location/Qualifiers
     source          1. .223156
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-275P22"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .5792
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    5893. .6990
                     /note="assembly_fragment"
     misc_feature    7091. .12953
                     /note="assembly_fragment"
     misc_feature    13054. .34245
                     /note="assembly_fragment"
     misc_feature    34346. .60533
                     /note="assembly_fragment"
     misc_feature    60634. .100582
                     /note="assembly_fragment"
     misc_feature    100683. .192104
                     /note="assembly_fragment"
     misc_feature    192205. .223156
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    67283 a  45092 c  44955 g  65105 t    721 others
ORIGIN
```

```
  Query Match            17.8%;  Score 178.6;  DB 2;  Length 223156;
  Best Local Similarity  69.1%;  Pred. No. 1e-19;
  Matches  244;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy        221 CTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGC 280
              |||| ||| |||| ||| || ||||||||      |   |    || ||| |||| ||| ||
Db     150068 CTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCGTGCTTGCTTGC 150127

Qy        281 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGC 340
              ||| ||||      |   |   || ||||||||      |   |    || ||| |||| ||
Db     150128 TTGCGTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCGTGCTTGC 150187

Qy        341 TGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGC 400
              | || ||||||||   | |   |   |  |||| |||||||  |   |  |||| ||| ||
Db     150188 TTGCTTGCTTGCTTGCGTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 150247

Qy        401 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 460
              ||| ||||||||  ||| || ||||||||||||  ||| || ||||||||||||  ||| ||
Db     150248 GTGCGTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 150307

Qy        461 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 520
              ||||||||||||  ||| || ||||||||||||  ||| || | | ||||||||  ||| ||
Db     150308 GTGCTTGCTTGCTTGCTTGCGTGCTTGCTTGCTTGCTTGCTTACGTGCTTGCTTGCTTGC 150367

Qy        521 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCT 573
              ||||||||||||  ||| || ||||||||      |   |   |  ||||||||
Db     150368 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCGTGCTTGCT 150420
```

RESULT 37
AC131371/c

```
LOCUS       AC131371               203152 bp    DNA     linear   HTG 15-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-318L15, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC131371
VERSION     AC131371.4  GI:25007315
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 203152)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 203152)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 203152)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 15, 2002 this sequence version replaced gi:23908406.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KCBL
                Center clone name: CH230-318L15
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 189099 bases at least Q40
                Consensus quality: 191887 bases at least Q30
                Consensus quality: 193446 bases at least Q20
                Estimated insert size: 194163; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   200227: contig of 200227 bp in length
```

```
          *   200228   200327: gap of unknown length
          *   200328   203152: contig of 2825 bp in length.
FEATURES             Location/Qualifiers
     source          1. .203152
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-318L15"
     misc_feature    1. .1183
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    2461. .4356
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    5237. .5898
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BZ141702"
     misc_feature    complement(198163. .199006)
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BZ141701"
BASE COUNT    47960 a  45439 c  46827 g  54363 t   8563 others
ORIGIN

  Query Match             17.6%;  Score 175.8;  DB 2;  Length 203152;
  Best Local Similarity   58.0%;  Pred. No. 3e-19;
  Matches  368;  Conservative    0;  Mismatches  262;  Indels    5;  Gaps    3;

Qy        219 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 278
              | |||| | | |||| | | || ||| |||| |  | | |  || |||||| | |||
Db     189394 GCCTGCTTCCTTGCTTGTTTGCTTGCCTGCTTGCTTGCCTGGCTGCTTGCTTGTTTGCTT 189335

Qy        279 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 338
              |||||| ||||     |   | | || ||||||||     | | |   || |||||| |
Db     189334 GCCTGCCTGCTTGCTTGTTTGAATGCTTGCTTGCTTGCCTGCCTGCCTGCTTGCTTGTTT 189275

Qy        339 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 398
              |   || ||||||||     | | |   |  || | |||||||     |   ||| |||
Db     189274 GAATGCTTGCTTGCTTGCCTGCCTGCCTGCTTGTTCGCTTGCTTACTTGCTCGCTTGCTC 189215

Qy        399 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 458
              || |||||||||||  ||| |  | || ||||||||||    |||||||||||||||
Db     189214 GCTTGCTTGCTTGCATGCTTGATTCCTGGCTTGCTCGCTTAAATGCTTGCTTGCTCGCTT 189155

Qy        459 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 518
              || |||||||||   | |  ||||||||||| ||  |||||  | | |||| |||
Db     189154 AAATGATTGCTTGCTTATTTGAATGCTTGCTTGCTTGCCTGCCTGACTACCTGCTTGCTT 189095

Qy        519 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTA 578
              ||    |||||||||| || || ||||| ||      |   |      ||||||||||||
Db     189094 GC---TTTGCTTGCTCACTTGCTTGCTTCCTTGCTTGCTTGCTTAAATGCTTGCTTGCTC 189038

Qy        579 AGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 638
              |   |   |   |  ||  || | ||||| | |   | ||||  | ||||  |   | ||
Db     189037 ACTTGCTTGCTTGTTTGAATGCTTGCTTGCTCGCTTGCCTGCCTGCCTGCCTGCCTGCCT 188978

Qy        639 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT-AGTAGGGC 697
              ||| |  ||||     |  |||| |  |||| |      |||| |  |||| | | |
Db     188977 GCTTGCTTGCTCACTTGCTTGCTTGAATGCTTGCTTCCTTGCTTGCTTGCTCACTTGCTT 188918

Qy        698 TCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGC 757
              | ||   |||| |   |  ||||       |     |  | || |  |||| |   |
Db     188917 GCTTGCTTTGCTTGCTTGCTTGCTCACTCAGTTCCTTGCTTGCTTGAATGCTTGCTTGCT 188858

Qy        758 TGCTGGGCTG-CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 816
              ||||  |||  || |   |    || |  || |     |  |||| |  ||||     |
Db     188857 TGCTTTGCTCACTTGCTTGCTCACTTGATTGTTCCCTTGCTTGCTTGTTTGCTCACCTGC 188798

Qy        817 CTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGG 851
                 ||| ||||  ||  ||||||  |||||| | |||

Db     188797 TTGCTTGCTCACTCACTTGCTCACTTGCTTGGGGG 188763


RESULT 38
AC095180/c
LOCUS       AC095180               270245 bp    DNA     linear   HTG 13-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-9C14, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC095180
VERSION     AC095180.5  GI:24940977
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 270245)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 270245)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
```

```
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 270245)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 13, 2002 this sequence version replaced gi:22772773.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GDAH
                Center clone name: CH230-9C14
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 230636 bases at least Q40
                Consensus quality: 232982 bases at least Q30
                Consensus quality: 234595 bases at least Q20
                Estimated insert size: 238600; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    74405: contig of 74405 bp in length
            *   74406    74505: gap of unknown length
            *   74506   196450: contig of 121945 bp in length
            *  196451   196550: gap of unknown length
            *  196551   270245: contig of 73695 bp in length.
FEATURES             Location/Qualifiers
     source          1. .270245
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-9C14"
     misc_feature    194462. .196450
                     /note="wgs_contig"
BASE COUNT    60901 a  56122 c  55375 g  63441 t  34406 others
ORIGIN

  Query Match             17.6%;  Score 175.8;  DB 2;  Length 270245;
  Best Local Similarity   58.0%;  Pred. No. 2.7e-19;
  Matches  368;  Conservative    0;  Mismatches  262;  Indels    5;  Gaps    3;

Qy        219 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 278
              | |||| | | |||| | | || ||| ||||     | | |   || |||||| | |||
Db     219492 GCCTGCTTCCTTGCTTGTTTGCTTGCCTGCTTGCTTGCCTGGCTGCTTGCTTGTTTGCTT 219433

Qy        279 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 338
              |||||| ||||     |   | | || ||||||||     | | |   || |||||| |
Db     219432 GCCTGCCTGCTTGCTTGTTTGAATGCTTGCTTGCTTGCCTGCCTGCCTGCTTGCTTGTTT 219373

Qy        339 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 398
              |   || ||||||||     | | |   |  || | |||||||     |   ||| |||
Db     219372 GAATGCTTGCTTGCTTGCCTGCCTGCCTGCTTGTTCGCTTGCTTACTTGCTCGCTTGCTC 219313

Qy        399 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 458
              || |||||||||||  ||| |  |. || ||||||||||    ||||||||||||||||
Db     219312 GCTTGCTTGCTTGCATGCTTGATTCCTGGCTTGCTCGCTTAAATGCTTGCTTGCTCGCTT 219253

Qy        459 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 518
              || |||||||||    | |  ||||||||||||| ||  |||||  | | |||| |||
Db     219252 AAATGATTGCTTGCTTATTTGAATGCTTGCTTGCTTGCCTGCCTGACTACCTGCTTGCTT 219193

Qy        519 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTA 578
              ||    |||||||||| || || ||||| ||      |   |      ||||||||||||
Db     219192 GC---TTTGCTTGCTCACTTGCTTGCTTCCTTGCTTGCTTGCTTAAATGCTTGCTTGCTC 219136

Qy        579 AGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 638
              |   |   |   |  ||  || | ||||| | |   | ||||  | ||||  |   | ||
Db     219135 ACTTGCTTGCTTGTTTGAATGCTTGCTTGCTCGCTTGCCTGCCTGCCTGCCTGCCTGCCT 219076

Qy        639 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT-AGTAGGGC 697
              . ||| |  ||||     |  |||| |  |||| |       |||| |  |||| | | |
Db     219075 GCTTGCTTGCTCACTTGCTTGCTTGAATGCTTGCTTCCTTGCTTGCTTGCTCACTTGCTT 219016

Qy        698 TCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGC 757
              | ||  |||| |   |  ||||       |     |  | || |  |||| |   |
Db     219015 GCTTGCTTTGCTTGCTTGCTTGCTCACTCAGTTCCTTGCTTGCTTGAATGCTTGCTTGCT 218956

Qy        758 TGCTGGGCTG-CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 816
              ||||  |||  || |   |    || |  || |     |  |||| |  ||||     |
Db     218955 TGCTTTGCTCACTTGCTTGCTCACTTGATTGTTCCCTTGCTTGCTTGTTTGCTCACCTGC 218896

Qy        817 CTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGG 851
              ||| ||||  ||  ||||||  |||||| | |||
Db     218895 TTGCTTGCTCACTCACTTGCTCACTTGCTTGGGGG 218861


RESULT 39
AC141889
LOCUS       AC141889              201258 bp    DNA     linear   HTG 20-MAR-2003
DEFINITION  Mus musculus chromosome UNK clone RP23-434P15, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC141889
VERSION     AC141889.2  GI:29126556
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 201258)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 201258)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 201258)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Mar 20, 2003 this sequence version replaced gi:29124296.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------- Project Information ----------
            Center project name: M_BA0434P15
```

```
          ----------------------------------------------------
          ------------- Summary Statistics  --------------
          Sequencing vector: M13; 0%
          Sequencing vector: plasmid; 100%
          Chemistry: Dye-primer ET; 0% of reads
          Chemistry: Dye-terminator Big Dye; 100% of reads
          Assembly program: Phrap; version 0.990319
          Consensus quality: 199803 bases at least Q40
          Consensus quality: 200560 bases at least Q30
          Consensus quality: 201124 bases at least Q20
          ----------------------------------------------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 9 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     1386: contig of 1386 bp in length
          *     1387     1486: gap of unknown length
          *     1487     3048: contig of 1562 bp in length
          *     3049     3148: gap of unknown length
          *     3149     6869: contig of 3721 bp in length
          *     6870     6969: gap of unknown length
          *     6970    26526: contig of 19557 bp in length
          *    26527    26626: gap of unknown length
          *    26627    52028: contig of 25402 bp in length
          *    52029    52128: gap of unknown length
          *    52129    77727: contig of 25599 bp in length
          *    77728    77827: gap of unknown length
          *    77828   109907: contig of 32080 bp in length
          *   109908   110007: gap of unknown length
          *   110008   153565: contig of 43558 bp in length
          *   153566   153665: gap of unknown length
          *   153666   201258: contig of 47593 bp in length.
FEATURES             Location/Qualifiers
     source          1. .201258
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-434P15"
     misc_feature    1. .1386
                     /note="assembly_name:Contig17"
     misc_feature    1487. .3048
                     /note="assembly_name:Contig19"
     misc_feature    3149. .6869
                     /note="assembly_name:Contig20"
     misc_feature    6970. .26526
                     /note="assembly_name:Contig21"
     misc_feature    26627. .52028
                     /note="assembly_name:Contig22"
     misc_feature    52129. .77727
                     /note="assembly_name:Contig23"
     misc_feature    77828. .109907
                     /note="assembly_name:Contig24"
     misc_feature    110008. .153565
                     /note="assembly_name:Contig25"
     misc_feature    153666. .201258
                     /note="assembly_name:Contig26"
BASE COUNT    54579 a  44579 c  45083 g  56182 t    835 others
ORIGIN

  Query Match             17.5%;  Score 175.2;  DB 2;  Length 201258;
  Best Local Similarity   56.2%;  Pred. No. 3.7e-19;
  Matches  370;  Conservative    0;  Mismatches  283;  Indels    5;  Gaps    2;

Qy        231 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 290
              ||| | ||||  | | | ||   ||   || ||   | | | | | ||||  | | | ||
Db      51183 GCTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTG 51242

Qy        291 AGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTT 350
              ||   || |||  | | | ||   ||   || ||   | | | | | |||   | |
Db      51243 GTTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTG 51302

Qy        351 GCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTT 410
              ||||   ||   |  ||   | ||| | | | || ||   | | ||||   ||| | |
Db      51303 GCTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGCTGGTTG 51362

Qy        411 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 470
              | | |||||   | | ||| | | ||||| ||| ||| ||| | | | |  || ||| ||
Db      51363 GTTGGCTGGTTGGTTGGCTGGTTGGCTGG-CTGGTTGGTTGGTTGGTTGGTTGGTT 51421

Qy        471 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 530
              | | | | | ||| |||  || | | | | ||| ||| |||   | | |  ||  || ||
Db      51422 GGTTGGTTGGCTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGCTGGTT 51481

Qy        531 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTAC 590
              | | | | | ||| ||| |  |  |   |   || || ||| ||| |  |  | | | |
Db      51482 GGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGGTT 51541

Qy        591 GCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 650
              | |||  ||   | ||| | |   || || | ||||| | |   ||||| | ||||| |
Db      51542 GGCTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGCTGGTT 51601

Qy        651 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTA 710
              |   ||||| | || || |     || || | ||||| | |   ||||  | || || |
Db      51602 GGTTGGCTGGTTGGTTGGTT----GGTTGGTTGGCTGGTTGGCTGGCTGGTTGGTTGGTT 51657

Qy        711 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 770
              |   || || | || || | |   ||||  | ||||| | |   || || | ||||| |
Db      51658 GGTTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGCTGGTTGGTTGGTTGGTTGGCTGGTT 51717

Qy        771 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTT 830
              |   ||||| | || ||   |   || || | ||||| | |   || ||  || | | ||
Db      51718 GGTTGGCTGGTTGGTTGGCTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGGTT 51777

Qy        831 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGG 888
              | ||| ||| ||| | |   ||| | ||| ||| |     || || |  || | | ||
Db      51778 GGTTGGTTGGTTGGTTGGCTGGCTGGTTGGCTGGTTGGCTGGTTGGTTGGCTGGTTGG 51835


RESULT 40
AF013613
LOCUS       AF013613               28199 bp    DNA     linear   VRT 21-NOV-1997
DEFINITION  Fugu rubripes cosmid 48D10, complete sequence.
ACCESSION   AF013613
VERSION     AF013613.1  GI:2627437
KEYWORDS    HTG.
SOURCE      Takifugu rubripes
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 28199)
  AUTHORS   Sandford,R., Sgotto,B., Aparicio,S., Brenner,S., Vaudin,M.,
            Wilson,R., Chissoe,S., Pepin,K., Bateman,A., Chothia,C., Hughes,J.
            and Harris,P.
  TITLE     Comparative analysis of the polycystic kidney disease 1 (PKD1) gene
            reveals an integral membrane glycoprotein with multiple
            evolutionary conserved domains
  JOURNAL   Hum. Mol. Genet. 6 (9), 1483-1489 (1997)
  MEDLINE   97449170
   PUBMED   9285785
REFERENCE   2  (bases 1 to 28199)
  AUTHORS   Hawkins,J. and Gillam,B.
  TITLE     The sequence of Fugu rubripes 48D10
  JOURNAL   Unpublished (1997)
REFERENCE   3  (bases 1 to 28199)
  AUTHORS   Washington University Genome Sequencing Center.
  TITLE     A Fugu rubripes Genome Sequencing Project
  JOURNAL   Unpublished (1997)
REFERENCE   4  (bases 1 to 28199)
  AUTHORS   Waterston,R.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Nov 21, 1997 this sequence version replaced gi:2623547.
            Submitted by:
                    Genome Sequencing Center
                    Department of Genetics, Washington University,
                    St. Louis, MO 63108, USA
                    e-mail: rw@nematode.wustl.edu


            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone

            Library information:
            Gridded Fugu library from the UK HGMP Resource Centre at Hinxton.


                        NEIGHBORING COSMID INFORMATION:


            The left neighbor is 259C6, 19100 base overlap. The actual start of
            this clone is at base position 22192 of 259C6; the actual end is at
            base position 28199 of 48D10.

            This cosmid contains a tandem repeat from 23145 to 25326. The
            tandem is deleted, and has regions which are single stranded.
            Attempts to size the deletion with digest data were inconclusive.
            The size was estimated to be
            1.7 kb from a PCR product obtained by Dr. Dick Sandford.

            NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES                Location/Qualifiers
     source             1. .28199
                        /organism="Takifugu rubripes"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:31033"
                        /chromosome="UL"
                        /map="unknown"
                        /clone="48D10"
     repeat_region      23145. .25326
                        /note="tandem with deletion"
     gene               26058. .28156
                        /gene="F_48D10.1"
     CDS                join(26058. .26702,27932. .28156)
                        /gene="F_48D10.1"
                        /note="Belongs to family 1 of G-protein coupled
                        receptors."
                        /codon_start=1
                        /evidence=not_experimental
                        /protein_id="AAB86684.1"
                        /db_xref="GI:2627438"
                        /translation="MEPLDQTPGFPLSPEPNYWYETTPSLLLVSYPHLLDISSNQSTQ
                        SVPFQGSSALLTAVIYITVFVVGLTGNTLAIYVVLRYAGMKTVTNIYILNLAVADELY
                        IVGLPFLATQNVLSYWPFGSFLCRVVMTADSMNQFTSIFCLTVMSIDRYLAVVHPIRS
                        TKWRHPRVAKVVSAAVWAVSFVVVLPVVIFSDVQVRPSRPLQVGTSSKCLVKRVQETF
                        NSCNMIWPEPKNVWSTAFILYTAMVGFFGPLLIICLCYLLIVIKVRHRMSAAQVGAVV
                        STCPLNICCLSRR"
BASE COUNT     7492 a    6758 c    6389 g    7560 t
ORIGIN

  Query Match             17.4%;  Score 174;  DB 5;  Length 28199;
  Best Local Similarity   53.4%;  Pred. No. 9.5e-19;
  Matches  366;  Conservative    0;  Mismatches  320;  Indels    0;  Gaps    0;

Qy        191 GTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTG 250
              || ||| ||  ||| | |  |  |  | ||| |||||||   ||  | |||  ||
Db      23583 GTCTGCCTGTCTGTCTGTCTGCCTGCCTGTCTGTCTGCCTGTCTGCCTGTCTGTCTGTCT 23642

Qy        251 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 310
              | | ||  | |||| ||   |   | |||| |||       | | ||  | |||| |
Db      23643 GCCTGCCTGTCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGCCTGCCTGTCTGTCTGCCT 23702

Qy        311 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 370
              |       | | ||  ||||| ||   |  |||||| |||  |   | | ||  | ||
Db      23703 GTCTGTCTGCCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGCCTACCTGCCTGTCTGTCT 23762

Qy        371 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 430
              ||  |  |||  ||  | ||||  |   | |||  |||  ||  |   | |||| ||  |
Db      23763 GCCTGTCTGCCCGTCTGTCTGCCTGTCTGTCTGTCTGCCCGCCTGTCTGTCTGCCTGTCT 23822

Qy        431 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              |   ||  | |||  ||| |||  ||  | |||| ||  ||   |   |||||  ||  |
Db      23823 GTCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGTCTGCCTGTCTGTCT 23882

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              ||  ||  |||||  ||  |||  |   | ||||  |  ||   ||  | |||  |||
Db      23883 GCCTGCCTGCCTGTCTGTCTGCCTGTCTGTCTGCCCGTCTGTCTGCCTGTCTGTCTGCCT 23942

Qy        551 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 610
              | | ||  | |||  ||  |||       | | |   |||||  |||   |  ||
Db      23943 GCCTGTCTGTCTGCCTGTCTGTCTGCCTGTCTGTCTGCCTGCCTGTCTGTCTGTCTGTCT 24002

Qy        611 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 670
              |   | ||||  | ||||  ||  | ||||  | ||||  ||  | |||   | |||
Db      24003 GCCTGCCTGCCTGCCTGCCTGTCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGTCTGTCT 24062

Qy        671 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 730
              |   | ||||  | |||   |   | || |  | ||||  |   | ||||  | |||
Db      24063 GCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGTCT 24122

Qy        731 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 790
              ||  | || |  | ||||  |   | |||   | ||||  ||  | ||||  | |||
Db      24123 GTCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTGCCTGTCTGTCTGCCTGTCTGTCTGTCT 24182

Qy        791 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGG 850
              ||  | ||||  | ||||  ||  | ||||||||  || ||| ||| ||  ||   ||
Db      24183 GTCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTGTCT 24242

Qy        851 GGCCGCTTGCCTGCTACTAGGGCTGC 876
              | | || |||||||   |  | ||||
Db      24243 GCCTGCCTGCCTGCCTGTCTGTCTGC 24268


RESULT 41
AX652127
LOCUS       AX652127               28199 bp    DNA     linear   PAT 22-MAR-2003
DEFINITION  Sequence 1978 from Patent WO03000898.
ACCESSION   AX652127
VERSION     AX652127.1  GI:29154941
KEYWORDS    .
SOURCE      Takifugu rubripes (Fugu rubripes)
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Takifugu.
REFERENCE   1
  AUTHORS   Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
            Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
  TITLE     Plant genes involved in defense against pathogens
  JOURNAL   Patent: WO 03000898-A 1978 03-JAN-2003;
            Syngenta Participations AG (CH)
FEATURES                Location/Qualifiers
     source             1. .28199
```

```
                     /organism="Takifugu rubripes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:31033"
BASE COUNT     7492 a   6758 c   6389 g   7560 t
ORIGIN

  Query Match           17.4%;  Score 174;  DB 6;  Length 28199;
  Best Local Similarity   53.4%;  Pred. No. 9.5e-19;
  Matches  366;  Conservative    0;  Mismatches  320;  Indels    0;  Gaps    0;

Qy        191 GTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTG 250
              || ||| ||  |||  | |  |  |   | ||| |||||||   ||  | |||  ||
Db      23583 GTCTGCCTGTCTGTCTGTCTGCCTGCCTGTCTGTCTGCCTGTCTGCCTGTCTGTCTGTCT 23642

Qy        251 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 310
              | | ||  | |||| ||   |   | |||| |||      | | ||  | |||| |
Db      23643 GCCTGCCTGTCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGCCTGCCTGTCTGTCTGCCT 23702

Qy        311 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 370
              |       | | ||  |||||  ||   |   ||||||| |||  |   | | ||  | ||
Db      23703 GTCTGTCTGCCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGCCTACCTGCCTGTCTGTCT 23762

Qy        371 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 430
              ||  |  |||  ||  | ||||  |   | |||  |||  ||  |   | ||||  ||  |
Db      23763 GCCTGTCTGCCCGTCTGTCTGCCTGTCTGTCTGTCTGCCCGCCTGTCTGTCTGCCTGTCT 23822

Qy        431 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              |   ||  | |||  ||| |||  ||  | |||| ||  ||   |   |||||  ||  |
Db      23823 GTCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGTCTGCCTGTCTGTCT 23882

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              ||  ||  |||||  ||  |||  |   | ||||  |  ||   ||  |  |||  |||
Db      23883 GCCTGCCTGCCTGTCTGTCTGCCTGTCTGTCTGCCCGTCTGTCTGCCTGTCTGTCTGCCT 23942

Qy        551 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 610
              | | ||  | |||  ||  |||  -    | | |   |||||  |||   |  ||
Db      23943 GCCTGTCTGTCTGCCTGTCTGTCTGCCTGTCTGTCTGCCTGCCTGTCTGTCTGTCTGTCT 24002

Qy        611 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 670
              |   | ||||  | ||||  ||  | ||||  | ||||  ||  | |||   | |||
Db      24003 GCCTGCCTGCCTGCCTGCCTGTCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGTCTGTCT 24062

Qy        671 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 730
              |   | ||||  | |||   |   | || |  | ||||  |   | ||||  | |||
Db      24063 GCCTGCCTGCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGTCT 24122

Qy        731 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 790
              ||  | || |  | ||||  |   | |||   | ||||  ||  | ||||  | |||
Db      24123 GTCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTGCCTGTCTGTCTGCCTGTCTGTCT 24182

Qy        791 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGG 850
              ||  | ||||  | ||||  ||  | ||||||||  || ||| ||| ||  ||   ||
Db      24183 GTCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTGTCT 24242

Qy        851 GGCCGCTTGCCTGCTACTAGGGCTGC 876
              | | || |||||||   |  | ||||
Db      24243 GCCTGCCTGCCTGCCTGTCTGTCTGC 24268


RESULT 42
AF388526
LOCUS       AF388526                599 bp    DNA     linear   PRI 16-AUG-2001
DEFINITION  Homo sapiens ZNF9 gene, intron 1 and expanded CL3N58 repeat region.
ACCESSION   AF388526
VERSION     AF388526.1  GI:15193054
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 599)
  AUTHORS   Liquori,C.L., Ricker,K., Moseley,M.L., Jacobsen,J.F., Kress,W.,
            Naylor,S.L., Day,J.W. and Ranum,L.P.
  TITLE     Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of
            ZNF9
  JOURNAL   Science 293 (5531), 864-867 (2001)
  MEDLINE   21378719
   PUBMED   11486088
REFERENCE   2  (bases 1 to 599)
  AUTHORS   Liquori,C.L., Jacobsen,J.F. and Ranum,L.P.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-2001) Genetics, Cell Biology, and Development,
            University of Minnesota, 516 Delaware St. SE, Minneapolis, MN
            55455, USA
FEATURES             Location/Qualifiers
     source          1. .599
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3q21; between D3S1589 and D3S3606"
     gene            <1. .>599
                     /gene="ZNF9"
                     /note="synonym: DM2; myotonic dystrophy type 2"
     intron          <1. .>599
                     /gene="ZNF9"
                     /number=1
     repeat_region   1. .599
                     /note="CL3N58 repeat region"
     repeat_region   69. .100
                     /rpt_type=tandem
                     /rpt_unit=tg
     repeat_region   101. .136
                     /rpt_type=tandem
                     /rpt_unit=tctg
     repeat_region   137. .576
                     /note="expanded repeat in DM2 patients"
                     /rpt_type=tandem
                     /rpt_unit=cctg
BASE COUNT       22 a    248 c    156 g    173 t
ORIGIN

  Query Match           17.3%;  Score 173.4;  DB 9;  Length 599;
  Best Local Similarity   58.7%;  Pred. No. 3.2e-18;
  Matches  300;  Conservative    0;  Mismatches  211;  Indels    0;  Gaps    0;

Qy        189 GCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGC 248
              | ||||| |||  |||| |     | ||  | ||| ||| |||   |   | |||  ||
Db         74 GTGTGTGTGTGTGTGTGTGTGTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 133

Qy        249 TGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGC 308
                  | |  |||||| |||  ||  |||||| |||      | | |   ||||||
Db        134 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGC 193

Qy        309 TTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGG 368
               |||      | | |  |||||| |||  ||. ||||||| |||      | | |   |
Db        194 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGC 253

Qy        369 CTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 428
              ||||  || |||  |   | ||||  ||  |||||| ||| |||  ||  |||||| |||
Db        254 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGC 313

Qy        429 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 488
               |||  ||  |||||| ||| |||  ||  |||||| ||| |||  ||  |||||| |||
Db        314 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGC 373

Qy        489 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 548
               |||  ||  |||||| ||| |||  ||  |||||| ||| |||  ||  |||||| |||
Db        374 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGC 433

Qy        549 TGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGA 608
                  | | |   | |||| ||| |||      | | |   |||||||||   || ||
Db        434 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGC 493

Qy        609 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGC 668
                |   | ||||  | ||||  |   | ||||  | ||||  |   | ||||  | ||||
```

```
Db        494 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGC 553

Qy        669 TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTC 699
              |   | ||||  | ||||  ||  | |||
Db        554 CTGCCTGCCTGCCTGCCTGTCTGTCTC 584


RESULT 43
AC091372
LOCUS       AC091372               235808 bp    DNA     linear   HTG 09-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-1C1, *** SEQUENCING IN PROGRESS ***,
            4 unordered pieces.
ACCESSION   AC091372
VERSION     AC091372.8  GI:22855787
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 235808)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 235808)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 235808)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 14, 2002 this sequence version replaced gi:22094155.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequening reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/).  As a result, the
            sequence may extend beyond the ends of the clone and there may be
            contigs that consist entirely of whole genome shotgun sequence
            reads.  Both end sequences and whole genome shotgun sequence only
            contigs will be indicated in the feature table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: TUFN
                Center clone name: CH230-1C1
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 222117 bases at least Q40
                Consensus quality: 223886 bases at least Q30
                Consensus quality: 224946 bases at least Q20
                Estimated insert size: 242733; sum-of-contigs estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   232025: contig of 232025 bp in length
            *    232026   232125: gap of unknown length
            *    232126   233295: contig of 1170 bp in length
            *    233296   233395: gap of unknown length
            *    233396   234502: contig of 1107 bp in length
            *    234503   234602: gap of unknown length
            *    234603   235808: contig of 1206 bp in length.
FEATURES             Location/Qualifiers
     source          1. .235808
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-1C1"
     misc_feature    1. .1198
                     /note="wgs_contig"
     misc_feature    1541. .2900
                     /note="wgs_contig"
BASE COUNT    67252 a  46511 c  47166 g  65258 t   9621 others
ORIGIN

  Query Match             17.3%;  Score 173.2;  DB 2;  Length 235808;
  Best Local Similarity   55.2%;  Pred. No. 7.4e-19;
  Matches  359;  Conservative    0;  Mismatches  288;  Indels    3;  Gaps    1;

Qy        221 CTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGC 280
              ||||||| |||   ||  ||||||  |||      | | |   | |||  |||  ||  ||
Db     228533 CTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTGCCTGC 228592

Qy        281 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGC 340
              ||||  |||      | | ||  ||||||  |||      | | ||  ||||||  ||   ||
Db     228593 CTGCCTGCCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGTCTGTCTGCCTGCCTGTCTGC 228652

Qy        341 TGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGC 400
                | ||||  |       | | |   | ||||     |||  ||  | |||   |   ||
Db     228653 CTGTCTGCCGGTCTGTCTGCCTGCCTGCCTGCCTTTCTGCCTGTCTGCCTGTCTGTCTGC 228712

Qy        401 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 460
```

```
Db     228713 CTGCCTGCCTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCTGACTGCC 228772

Qy        461 CTGCTTGCTTGCTCGCTG---GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 517

Db     228773 TGCCTTTCTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTTTCTGCCTGTCTGTC 228832

Qy        518 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCT 577

Db     228833 TGCCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCC 228892

Qy        578 AAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGC 637

Db     228893 TTTCTGCCTGTCTGTCTGCCTACCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTGCCTGCC 228952

Qy        638 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 697

Db     228953 TGCCTTTCTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTACC 229012

Qy        698 TCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGC 757

Db     229013 TGCCTGCCTGCCTGTCTGTCTGACTGACTGCCTGCCTGCCTGCCTGCCTTTCTGCC 229072

Qy        758 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 817

Db     229073 TGTCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTACCTGCCTGCCTGCC 229132

Qy        818 TGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTAC 867

Db     229133 TGTCTGTCTGACTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCTGC 229182
```

RESULT 44
AC107437/c
LOCUS AC107437 305226 bp DNA linear HTG 13-MAY-2003
DEFINITION Rattus norvegicus clone CH230-84A6, *** SEQUENCING IN PROGRESS ***, 25 unordered pieces.
ACCESSION AC107437
VERSION AC107437.6 GI:30580694
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 305226)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 305226)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 305226)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 13, 2003 this sequence version replaced gi:25188618.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
-------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
Center project name: GLDM
Center clone name: CH230-84A6
-------------- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 283306 bases at least Q40
Consensus quality: 287489 bases at least Q30
Consensus quality: 290737 bases at least Q20
Estimated insert size: 299662; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    11777: contig of 11777 bp in length
            *    11778    11877: gap of unknown length
            *    11878   258209: contig of 246332 bp in length
            *   258210   258309: gap of unknown length
            *   258310   259657: contig of 1348 bp in length
            *   259658   259757: gap of unknown length
            *   259758   261089: contig of 1332 bp in length
            *   261090   261189: gap of unknown length
            *   261190   262280: contig of 1091 bp in length
            *   262281   262380: gap of unknown length
            *   262381   264273: contig of 1893 bp in length
            *   264274   264373: gap of unknown length
            *   264374   265950: contig of 1577 bp in length
            *   265951   266050: gap of unknown length
            *   266051   267410: contig of 1360 bp in length
            *   267411   267510: gap of unknown length
            *   267511   268811: contig of 1301 bp in length
            *   268812   268911: gap of unknown length
            *   268912   270237: contig of 1326 bp in length
            *   270238   270337: gap of unknown length
            *   270338   271796: contig of 1459 bp in length
            *   271797   271896: gap of unknown length
            *   271897   274424: contig of 2528 bp in length
            *   274425   274524: gap of unknown length
            *   274525   275942: contig of 1418 bp in length
            *   275943   276042: gap of unknown length
            *   276043   277659: contig of 1617 bp in length
            *   277660   277759: gap of unknown length
            *   277760   279936: contig of 2177 bp in length
            *   279937   280036: gap of unknown length
            *   280037   281813: contig of 1777 bp in length
            *   281814   281913: gap of unknown length
            *   281914   283894: contig of 1981 bp in length
            *   283895   283994: gap of unknown length
            *   283995   286131: contig of 2137 bp in length
            *   286132   286231: gap of unknown length
            *   286232   287386: contig of 1155 bp in length
            *   287387   287486: gap of unknown length
            *   287487   289613: contig of 2127 bp in length
            *   289614   289713: gap of unknown length
            *   289714   291967: contig of 2254 bp in length
            *   291968   292067: gap of unknown length
            *   292068   293964: contig of 1897 bp in length
            *   293965   294064: gap of unknown length
            *   294065   295962: contig of 1898 bp in length
            *   295963   296062: gap of unknown length
            *   296063   300082: contig of 4020 bp in length
            *   300083   300182: gap of unknown length
            *   300183   305226: contig of 5044 bp in length.
FEATURES             Location/Qualifiers
     source          1. .305226
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-84A6"
     misc_feature    789. .37330
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH277447"
     misc_feature    10136. .11777
                     /note="wgs_contig"
     misc_feature    complement(239138. .240020)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH276631"
     misc_feature    253535. .254639
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    256559. .258209
                     /note="wgs_end_extension
                     clone_end:T7"
BASE COUNT    83055 a  60966 c  60899 g  88408 t  11898 others
ORIGIN

  Query Match             17.3%;  Score 173.2;  DB 2;  Length 305226;
  Best Local Similarity   55.2%;  Pred. No. 6.9e-19;
  Matches  359;  Conservative    0;  Mismatches  288;  Indels    3;  Gaps    1;

Qy        221 CTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGC 280
              ||||||| |||   ||  |||||| |||      | | |   | |||  |||  ||  ||
Db      88519 CTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTGCCTGC 88460

Qy        281 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGC 340
              ||||  |||      | | ||  |||||| |||      | | ||  |||||| ||   ||
Db      88459 CTGCCTGCCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGTCTGTCTGCCTGCCTGTCTGC 88400

Qy        341 TGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGC 400
                | ||||  |       | | |   | ||||     |||  ||  | |||   |   ||
Db      88399 CTGTCTGCCGGTCTGTCTGCCTGCCTGCCTGCCTTTCTGCCTGTCTGCCTGTCTGTCTGC 88340

Qy        401 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 460
              |||| ||| |||  |   | |||| ||| |||  ||  |||||  ||| ||||  ||| |
Db      88339 CTGCCTGCCTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCTGACTGCC 88280

Qy        461 CTGCTTGCTTGCTCGCTG---GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 517
                ||| ||  ||  |||   |||||| ||| |||  ||  ||||   ||| ||   |
Db      88279 TGCCTTTCTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGCCTGCCTTTCTGCCTGTCTGTC 88220

Qy        518 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCT 577
              ||||||  ||| ||   ||  | |||| |||      | | ||  | |||| ||| |||
Db      88219 TGCCTGCCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCC 88160

Qy        578 AAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGC 637
                | | ||  | ||||||    || ||   ||  | |||   | ||||  |   | |
Db      88159 TTTCTGCCTGTCTGTCTGCCTACCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTGCCTGCC 88100

Qy        638 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 697
              |||    ||||  ||  | ||||  | ||||  |   | ||||  | |||   |     |
Db      88099 TGCCTTTCTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTACC 88040

Qy        698 TCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGC 757
              | |  | |||  ||  | |||   | ||||  |   | || |  | ||||   |  | |
Db      88039 TGCCTGCCTGCCTGTCTGTCTGACTGACTGCCTGCCTGCCTGCCTGCCTTTCTGCC 87980

Qy        758 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 817
              ||   | ||||  |   | ||||  | ||||  ||  | ||||    ||||  |   | |
Db      87979 TGTCTGTCTGCCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTACCTGCCTGCCTGCC 87920

Qy        818 TGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTAC 867
              || |||   |  ||| ||| ||| |||  |   | | |  ||||||||
Db      87919 TGTCTGTCTGACTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCTGC 87870


RESULT 45
AC098644
LOCUS       AC098644               88716 bp    DNA     linear   HTG 30-AUG-2002
DEFINITION  Takifugu rubripes clone 205P11, WORKING DRAFT SEQUENCE, 7 ordered
            pieces.
ACCESSION   AC098644
VERSION     AC098644.3  GI:22549754
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Takifugu rubripes (Fugu rubripes)
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 88716)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
```

```
            Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
            Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Reddix-Dugue,N., Schueler,M.G., Sison,C.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 88716)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 88716)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-AUG-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Aug 30, 2002 this sequence version replaced gi:17155042.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: arp
                 Center clone name: 205P11

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 87141 bases at least Q40
                 Consensus quality: 87741 bases at least Q30
                 Consensus quality: 87993 bases at least Q20
                 Insert size: 89000; agarose-fp
                 Insert size: 88116; sum-of-contigs
                 Quality coverage: 10.68x in Q20 bases; agarose-fp
                 Quality coverage: 10.79x in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    31156: contig of 31156 bp in length
            *    31157    31256: gap of unknown length
            *    31257    36515: contig of 5259 bp in length
            *    36516    36615: gap of unknown length
            *    36616    38675: contig of 2060 bp in length
            *    38676    38775: gap of unknown length
            *    38776    41147: contig of 2372 bp in length
            *    41148    41247: gap of unknown length
            *    41248    48183: contig of 6936 bp in length
            *    48184    48283: gap of unknown length
            *    48284    50318: contig of 2035 bp in length
            *    50319    50418: gap of unknown length
            *    50419    88716: contig of 38298 bp in length.
FEATURES             Location/Qualifiers
     source          1. .88716
                     /organism="Takifugu rubripes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:31033"
                     /clone="205P11"
                     /clone_lib="Incyte Genomics"
     misc_feature    1. .31156
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    31257. .36515
                     /note="assembly_fragment"
     misc_feature    36616. .38675
                     /note="assembly_fragment"
     misc_feature    38776. .41147
                     /note="assembly_fragment"
     misc_feature    41248. .48183
                     /note="assembly_fragment"
     misc_feature    48284. .50318
                     /note="assembly_fragment"
     misc_feature    50419. .88716
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    52511. .88716
                     /note="clone overlaps with GenBank Accession Number
                     AC098643 clone 194A9 (center project name aro)"
BASE COUNT    23629 a  19624 c  20682 g  24180 t    601 others
ORIGIN

  Query Match           17.3%;  Score 173;  DB 2;  Length 88716;
  Best Local Similarity   53.3%;  Pred. No. 1e-18;
  Matches  365;  Conservative    0;  Mismatches  320;  Indels    0;  Gaps    0;

Qy        191 GTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTG 250
              || ||| ||| |||  | |  |  ||  | |||||||||||| ||   | |||  ||
Db      33854 GTCTGCCTGCCTGTCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGTCTGTCTGTCTGTCT 33913

Qy        251 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 310
                | | |    ||||  |||  |   |||||| ||       | | ||  |||||| |
Db      33914 GTCTGTCTGCCTACCTGTCTGCCTGTCTGCCTGCCTGTCTGTCTGTCTGTCTGCCTGCCT 33973

Qy        311 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 370
              |       | | ||  | |||| |||  ||  |||||| ||      | | |   | ||
Db      33974 GTCTGCCTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGTCT 34033

Qy        371 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 430
              |   || ||   ||  | ||||  ||  | |||| ||  ||   |  |||||| ||| |
Db      34034 GTCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGTCTGCCTGCCTGCCT 34093

Qy        431 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              | |  | |||||  ||| |||  ||  | |||| ||  ||   ||  |||||  ||  |
Db      34094 GTTTACCTGCCTGTCTGCCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTGCCTGTCTGTCT 34153

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              |   |   |||||| || || |  ||  | |||| ||| ||   |   | |||  ||
Db      34154 GTCTGTCTGCCTGCCTGTTTACCTGCCTGTCTGCCTGCCTGTCTGTCTGTCTGTCTGTCT 34213

Qy        551 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 610
                | | ||  | |||  ||| |||       | | |   | ||| |||   || ||
Db      34214 GTCTGTCTGTCTGTCTGTCTGCCTGCCTGCCTGCCTGTCTGTCTGTCTGCATGCCT 34273

Qy        611 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 670
              ||  | ||||  | ||||  ||  | ||||  | |||   ||  | |||   | |||
Db      34274 GTCTGTCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 34333

Qy        671 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 730
              |  | ||||  | ||||  |  | ||    | |||   ||  | ||||  | |||
Db      34334 GCCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTGTCTGTCTGTCTGTCTGCCTGCCTGTCT 34393
```

```
Qy        731 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 790
              ||  | ||    | |||   ||  | ||||  | |||   ||  | ||||  | |||
Db      34394 GTCTGTCTGTCTGTCTGTCTGTCTGCCTGCCTGCCTGTCTGTCTGTCTGCCTGCCTGTCT 34453

Qy        791 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGG 850
              |   | |||   | ||||  ||  | ||| |||   |  ||  ||| ||| ||   ||
Db      34454 GCCTGTCTGTCTGCCTGCCTGTCTGTCTGTCTGTCTGTCTGTCTGCCTGCCTGTCTGTCT 34513

Qy        851 GGCCGCTTGCCTGCTACTAGGGCTG 875
              | | || ||||||    |  | |||
Db      34514 GTCTGCCTGCCTGTCTGTCTGTCTG 34538


RESULT 46
AC109837/c
LOCUS       AC109837               197891 bp    DNA     linear   HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-271E13, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC109837
VERSION     AC109837.5  GI:24818656
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 197891)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 197891)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 197891)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 9, 2002 this sequence version replaced gi:23101414.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GOGT
                Center clone name: CH230-271E13
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 191506 bases at least Q40
                Consensus quality: 192743 bases at least Q30
                Consensus quality: 193813 bases at least Q20
                Estimated insert size: 195908; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   193483: contig of 193483 bp in length
            *  193484   193583: gap of unknown length
            *  193584   194691: contig of 1108 bp in length
            *  194692   194791: gap of unknown length
            *  194792   196008: contig of 1217 bp in length
            *  196009   196108: gap of unknown length
            *  196109   197891: contig of 1783 bp in length.
FEATURES             Location/Qualifiers
     source          1. .197891
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-271E13"
     misc_feature    1. .1079
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    complement(190998. .191740)
```

```
                     /note="clone_boundary
                     clone_end:Sp6
                     site:MboI
                     end_sequence:BZ218549"
     misc_feature    complement(191078. .191745)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:MboI
                     end_sequence:BZ218549"
BASE COUNT    62207 a  36875 c  35461 g  60011 t   3337 others
ORIGIN

  Query Match             17.2%;  Score 172.4;  DB 2;  Length 197891;
  Best Local Similarity   53.2%;  Pred. No. 1e-18;
  Matches  365;  Conservative    0;  Mismatches  321;  Indels    0;  Gaps    0;

Qy        194 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 253
              | | ||| ||   | |  |  |   | ||||||||||||  ||  |||||| |||
Db      19470 TCCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCC 19411

Qy        254 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 313
              | | ||| | |||| |     |  |||||| |||    | |        || |   |||
Db      19410 TGCCTGTATGTCTGCCTAACTCCCTGCCTGCCTGCCTGCGTGCATACGTGCGTCAGTGCG 19351

Qy        314 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 373
              | | | | || | | |||   |  |||||  ||        ||     | ||||
Db      19350 TGCAAGCCTGCATGCATACATGCATCCCTGCCTGTATGTCTGCATAACTCCTTGCCTGCC 19291

Qy        374 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 433
              || ||| |  | | | ||  ||  || ||| | | |||  ||  |  ||  ||| |
Db      19290 TGCCTGCCAAAAAGACAGCCTGCCTGCATGCCTACATGCGTGCCTGTATGTCTGCCTAAC 19231

Qy        434 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 493
              |  || ||| ||| |||  ||  || ||| ||| |||   |  |||||  ||  |||
Db      19230 TCCCTGCATGCATGCCTGCGTGCATGCATGCCTGCATGCCTACCTGCCTGTATGTCTGCA 19171

Qy        494 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGG 553
              | | | ||| |||  ||  |||||| ||| |||  ||  |||||| |||
Db      19170 TAACTCCATACATGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTA 19111

Qy        554 GGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTA 613
              | | |     || | ||| |||       | | |   |||||||||    | ||   |||
Db      19110 TGTCTGCCTAACTCCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCTGCCTGTA 19051

Qy        614 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 673
              | ||||    || |  |  | ||||  | ||||  |   |  | ||||  | ||||  |
Db      19050 TGTCTGCCTAACTCCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCC 18991

Qy        674 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 733
              | || |  | |||   ||  | ||      ||||  |   | ||||  | ||||  |
Db      18990 TGCCTCCCTGCCTGTATGTCTGCCTAACTCCCTGCCTGCCTGCCTGCCTGCCTGCC 18931

Qy        734 GGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 793
              | || |  | ||||  |   | ||||  | ||||  |   | ||||    || |  |
Db      18930 TGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTAACTCCCTGCC 18871

Qy        794 GGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGC 853
              | ||||  | ||||  |   | ||||||||  || ||| ||| ||| |||  |   | |
Db      18870 TGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCC 18811

Qy        854 CGCTTGCCTGCTACTAGGGCTGCTGT 879
              || | ||||    |  | |||||||
Db      18810 TGCCTCCCTGGCTGTCTGCCTGCTGT 18785


RESULT 47
AC113840/c
LOCUS       AC113840               245258 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-126F2, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC113840
VERSION     AC113840.5  GI:25072829
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 245258)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 245258)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 245258)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:23101014.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
```

in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GSOU
Center clone name: CH230-126F2
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 228855 bases at least Q40
Consensus quality: 231441 bases at least Q30
Consensus quality: 233224 bases at least Q20
Estimated insert size: 235507; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
---
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 78039: contig of 78039 bp in length
* 78040 78139: gap of unknown length
* 78140 241585: contig of 163446 bp in length
* 241586 241685: gap of unknown length
* 241686 245258: contig of 3573 bp in length.

```
FEATURES             Location/Qualifiers
     source          1. .245258
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-126F2"
     misc_feature    1683. .2513
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BH358726"
     misc_feature    6441. .7279
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BH358725"
     misc_feature    78140. .80241
                     /note="wgs_end_extension
                     clone_end:T7"
BASE COUNT    62337 a  53440 c  55280 g  63518 t  10683 others
ORIGIN
```

Query Match 16.9%; Score 169.4; DB 2; Length 245258;
Best Local Similarity 53.9%; Pred. No. 2.9e-18;
Matches 370; Conservative 0; Mismatches 316; Indels 1; Gaps 1;

```
Qy        190 CGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCT 249
              ||  ||| ||  |    | |  |  ||  | ||||||||||||  ||  |||||  |||
Db      98980 CGCCTGCCTGTCTCCCTGTCTGCCTGTCTGTCTGCCTGCCTGCCTGCCTGCCTGTCTGCC 98921

Qy        250 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCT 309
                | | ||  |||||  |||  |   |||||  |||      | | ||  |||||
Db      98920 TGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGTCTGCCTGTC 98861

Qy        310 TGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGC 369
              |||      | | ||  | |||| |||  ||  | |||| |||      | |  |  | |
Db      98860 TGCCTGTCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGCCTGCCTGTCTGCCTATCTGCC 98801

Qy        370 TGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 429
              |||  ||·|||  ||  | |||    |  ||||||| ||| |    ||  |||||| |||
Db      98800 TGCCTGCCTGCCTGTCTGCCTGTCTCCCTGCCTGCCTGCCTATCTGCCTGCCTGCCTGCC 98741

Qy        430 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 489
              |||  |    ||||| ||| |||      |||||  ||| ||   ||  |||||| |||
Db      98740 TGCCTGTCTCCCTGCCTGCCTGCCTATCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCC 98681

Qy        490 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              |||  |   |||||  ||| ||   |   |||||| ||| ||   |   | |||  | |
Db      98680 TGCCTGTCTGCCTGTCTGCCTGTCTGTCTGCCTGCCTGCCTGTCTGTCTGACTGTCTACC 98621

Qy        550 GAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGAT 609
                | | ||  | |||  ||| |||      ||| ||   |||| |||   || ||
Db      98620 TGTCTGTCTGTCTGCCTGTCTGCCTGCCTGTCTGACTGTCTACCTGTCTGCCTGCCTGCC 98561

Qy        610 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 669
              ||  | ||||  | || |  |   | ||||    ||||  ||  | |||  | ||||
Db      98560 TGTCTGCCTGCCTGTCTCCCTGCCTGCCTGCCTATCTGCCTGTCTGCCTGTCTGCCTGCC 98501

Qy        670 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 729
              |   | ||| |  || ||  ||  | || |  | ||||  |   | || |  | ||||
Db      98500 TGCCTGCCTG-TCTGCCGCCTGTCTGCCTGCCTGTCTGCCTGCCTGCCTACCTGTCTGCC 98442

Qy        730 AGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 789
              |  | ||    | ||    |  | ||||  | ||||  |     ||||  | ||||
Db      98441 TATCTGCCTATCTGCCTATCTGCCTGTCTGCCTGCCTGCCTGCCTATCTGCCTGCCTGCC 98382

Qy        790 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTG 849
              ||  | |||  | ||||  ||  | ||| |||   || ||  ||| ||  ||   |
Db      98381 TGTCTGTCTGACTGTCTGCCTGTCTGCCTGTCTGTCTGCCTGTCTGCCTGTCTGTCTGCC 98322

Qy        850 GGGCCGCTTGCCTGCTACTAGGGCTGC 876
              | | |  |||||||   |  | ||||
Db      98321 TGTCTGTCTGCCTGCCTGTCTGTCTGC 98295
```

```
RESULT 48
AC141889/c
LOCUS       AC141889              201258 bp    DNA     linear   HTG 20-MAR-2003
DEFINITION  Mus musculus chromosome UNK clone RP23-434P15, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC141889
VERSION     AC141889.2  GI:29126556
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 201258)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 201258)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 201258)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Mar 20, 2003 this sequence version replaced gi:29124296.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
```

```
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information ------------------
            Center project name: M_BA0434P15
            ----------------------------------------------------
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 199803 bases at least Q40
            Consensus quality: 200560 bases at least Q30
            Consensus quality: 201124 bases at least Q20
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1386: contig of 1386 bp in length
            *     1387     1486: gap of unknown length
            *     1487     3048: contig of 1562 bp in length
            *     3049     3148: gap of unknown length
            *     3149     6869: contig of 3721 bp in length
            *     6870     6969: gap of unknown length
            *     6970    26526: contig of 19557 bp in length
            *    26527    26626: gap of unknown length
            *    26627    52028: contig of 25402 bp in length
            *    52029    52128: gap of unknown length
            *    52129    77727: contig of 25599 bp in length
            *    77728    77827: gap of unknown length
            *    77828   109907: contig of 32080 bp in length
            *   109908   110007: gap of unknown length
            *   110008   153565: contig of 43558 bp in length
            *   153566   153665: gap of unknown length
            *   153666   201258: contig of 47593 bp in length.
FEATURES             Location/Qualifiers
     source          1. .201258
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-434P15"
     misc_feature    1. .1386
                     /note="assembly_name:Contig17"
     misc_feature    1487. .3048
                     /note="assembly_name:Contig19"
     misc_feature    3149. .6869
                     /note="assembly_name:Contig20"
     misc_feature    6970. .26526
                     /note="assembly_name:Contig21"
     misc_feature    26627. .52028
                     /note="assembly_name:Contig22"
     misc_feature    52129. .77727
                     /note="assembly_name:Contig23"
     misc_feature    77828. .109907
                     /note="assembly_name:Contig24"
     misc_feature    110008. .153565
                     /note="assembly_name:Contig25"
     misc_feature    153666. .201258
                     /note="assembly_name:Contig26"
BASE COUNT    54579 a  44579 c  45083 g  56182 t    835 others
ORIGIN

  Query Match             16.9%;  Score 169.2;  DB 2;  Length 201258;
  Best Local Similarity   54.1%;  Pred. No. 3.3e-18;
  Matches  367;  Conservative    0;  Mismatches  308;  Indels    3;  Gaps    1;

Qy        214 GGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 273
                  ||| |   |  ||   ||| |||||   | | ||||   ||   || ||   ||| |||
Db      76525 GGTTGGCTGGTTTGTTGGCTGGCTGGTTGGTTGGCTGGTTGGTTGGTTGGCTGGCT 76466

Qy        274 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCT 333
              |||||   ||| | ||   ||   || ||   | | | ||   ||   || ||| |||
Db      76465 GGCTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCT 76406

Qy        334 TGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCT 393
              ||| |||||   ||| | ||   ||   || ||   | ||| | ||| |  ||   | |
Db      76405 GGCTGGCTGGTTGGCTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGCTGGCTGGCTGGTT 76346

Qy        394 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC---CTGCTTGCTT 450
              | ||||  | | | | | | | |||   | | ||| ||| ||||||   ||| ||| ||
Db      76345 GGTTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGGCTGGCTGGCTGGTTGGTT 76286

Qy        451 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 510
              | | | | |  ||  || ||| | |    |  || |||  || | | | |  || ||| ||
Db      76285 GGTTGGTTGGTTGGCTGGTTGATTGGCTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTT 76226

Qy        511 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTT 570
              | | | | |  || ||| |||   | | |  || ||| |  |  | | |   || || ||
Db      76225 GGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTT 76166

Qy        571 GCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTA 630
              | ||| |  |      |   |  ||  || | |  || |||   || || | || ||
Db      76165 GGTTGGTTGGTTAGTTGGTTGGTTGGTTGGCTGGTAGGTTGGTTGGTTGGTTGGCT 76106

Qy        631 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 690
              |  ||||| | ||||   ||   || || | ||||| | |    | || | || || | |
Db      76105 GGTTGGCTGGTTGGCTAAAAGGTTGGTTGGTTGGCTGGTTGGTTAGTTGGTTGGATGGTT 76046

Qy        691 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTA 750
              |   || |  | |||||  | |   || ||    | |         ||||  | || || |
Db      76045 GGTTGGTTGGTTGGCTGGTTGGTTGGTTGGCTAGTTAGCTAGTTGGCTGGTTGGTTGGTT 75986

Qy        751 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 810
              |    |||| | ||||| | |   || || | || || | |   || || | ||||| |
Db      75985 GGTTAGCTGGTTGGCTGATTGGTTGGTTGATTGGTTGGTTGACTGGTTGGTTGGCTGGTT 75926

Qy        811 GTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAG 870
              |   || ||  || | | ||| ||| ||| ||| | |   ||  | |||  || |
Db      75925 GGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 75866

Qy        871 GGCTGCTGTGCAGCTGGG 888
              || || |  |  | | ||
Db      75865 GGTTGGTTGGTTGGTTGG 75848


RESULT 49
AC096407
LOCUS       AC096407               275443 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-59F13, *** SEQUENCING IN PROGRESS
            ***.
ACCESSION   AC096407
VERSION     AC096407.6  GI:30520472
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 275443)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
```

```
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 275443)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 275443)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24819680.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GFHU
                Center clone name: CH230-59F13
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 261063 bases at least Q40
                Consensus quality: 264086 bases at least Q30
                Consensus quality: 266112 bases at least Q20
                Estimated insert size: 272725; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *         1   275443: contig of 275443 bp in length.
FEATURES              Location/Qualifiers
     source           1. .275443
                      /organism="Rattus norvegicus"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:10116"
                      /clone="CH230-59F13"
     misc_feature     1. .1557
                      /note="wgs_end_extension
                      clone_end:Sp6"
     misc_feature     2530. .2683
                      /note="clone_boundary
                      clone_end:Sp6
                      site:EcoRI
                      end_sequence:BH332249"
     misc_feature     complement(266627. .267346)
                      /note="clone_boundary
                      clone_end:T7
                      site:EcoRI
                      end_sequence:BH332247"
     misc_feature     272350. .274094
                      /note="wgs_end_extension
                      clone_end:T7"
     misc_feature     274145. .275443
                      /note="wgs_end_extension
                      clone_end:T7"
BASE COUNT    80036 a   54787 c   53947 g   78751 t    7922 others
ORIGIN

  Query Match             16.7%;  Score 167.6;  DB 2;  Length 275443;
  Best Local Similarity   50.7%;  Pred. No. 5.5e-18;
  Matches  355;  Conservative    0;  Mismatches  344;  Indels    1;  Gaps    1;

Qy        176 GGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCG 235
              ||                                                       ||| |
Db      46322 GGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGGCTAG 46381

Qy        236 CTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 295
              |||||  | |   ||   ||   || ||   | | | | | ||||   || ||||   ||
Db      46382 CTGGCTGGTTGTTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGACTGGCTGGTTGG 46441

Qy        296 ACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGA 355
                || ||    || | ||   ||   || ||   | | | | ||||||  ||| ||||
Db      46442 TTGGTTGGTTGACTGGTTGGTTGGTTGGTTGGTTGGATGGTTGGCTGGCTGGCTGGCTGG 46501

Qy        356 GGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCG 415
                ||   || ||   | ||| | | | || ||   ||| |||| |   || | | | | |
Db      46502 CTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGACTGACTGGTTGGTTGG 46561

Qy        416 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 475
              |||    || | | | | | |||   | | | | | | ||||||  ||| ||| ||| |
Db      46562 TTGGTTGACTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGGCTGGCTGGCTGG 46621
```

```
Qy        476 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 535
              |||| || ||| ||| | | | | |||  ||  || | | | |  || ||| ||| | |
Db      46622 CTGG-TTGGTTGGTTGGTTGGCTGGCTGGCTGGTTGGTTGGTTGGTTGGTTG 46680

Qy        536 CTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTG 595
                | ||| ||| |  |  |   |  ||||| ||| ||| |  |  ||| |   |  ||
Db      46681 GCTGACTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGCTGACTGGCTGGTTG 46740

Qy        596 CCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGG 655
              || | | ||||  |   || || | || || | |   || || | |||||   |   |
Db      46741 GTTGGTTGGTTGACTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGGCTG 46800

Qy        656 GCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGG 715
              ||||  |||||   |   || || | || || | |   ||||   | |||   |   |
Db      46801 GCTGGCTGGCTGGCTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGACTGACTGGTTG 46860

Qy        716 GCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGG 775
              | || | || ||   |   ||||    || || | |   |||||  |||||   |   |
Db      46861 GTTGGTTGGTTGGCTGGCTGGCTGGCTGGTTGGTTGGTTGGCTGGCTGGCTGACTGGTTG 46920

Qy        776 GCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTG 835
              ||||  |||||   |   | ||| | || || | |   ||||| |||   | ||| |||
Db      46921 GCTGGCTGGCTGGCTGGCTGACTGGTTGGTTGGTTGGCTGGCTGGCTGACTGGTTGGTTG 46980

Qy        836 CTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTG 875
              |||  || | |   ||   |||  || |     |||||
Db      46981 GTTGGCTGGTTGGTTGGTTCGTTGGTTGGTTGGCTGGCTG 47020
```

RESULT 50
AC096597/c
LOCUS AC096597 271233 bp DNA linear HTG 13-NOV-2002
DEFINITION Rattus norvegicus clone CH230-21E2, *** SEQUENCING IN PROGRESS ***, 5 unordered pieces.
ACCESSION AC096597
VERSION AC096597.6 GI:24941421
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 271233)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 271233)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (19-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 271233)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 13, 2002 this sequence version replaced gi:22772261.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
-------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
Center project name: GEHU
Center clone name: CH230-21E2
-------------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 219869 bases at least Q40
Consensus quality: 225773 bases at least Q30
Consensus quality: 229378 bases at least Q20
Estimated insert size: 221309; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 10670: contig of 10670 bp in length

```
           *    10671    10770: gap of unknown length
           *    10771   266157: contig of 255387 bp in length
           *   266158   266257: gap of unknown length
           *   266258   267281: contig of 1024 bp in length
           *   267282   267381: gap of unknown length
           *   267382   268591: contig of 1210 bp in length
           *   268592   268691: gap of unknown length
           *   268692   271233: contig of 2542 bp in length.
FEATURES             Location/Qualifiers
     source          1. .271233
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-21E2"
     misc_feature    6418. .8047
                     /note="wgs_contig"
     misc_feature    49552. .51724
                     /note="wgs_contig"
     misc_feature    148748. .149962
                     /note="wgs_contig"
     misc_feature    150054. .151277
                     /note="wgs_contig"
     misc_feature    247156. .248507
                     /note="wgs_contig"
BASE COUNT    68348 a  46760 c  47960 g  69593 t  38572 others
ORIGIN

  Query Match             16.7%;  Score 167.4;  DB 2;  Length 271233;
  Best Local Similarity   54.7%;  Pred. No. 5.9e-18;
  Matches  353;  Conservative    0;  Mismatches  291;  Indels    1;  Gaps    1;

Qy        231 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 290
              ||| ||||||  | |   ||   ||   || ||   | | | | | ||||   || ||||
Db     221948 GCTAGCTGGCTGGTTGTTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGACTGGCTG 221889

Qy        291 AGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTT 350
                ||   || ||   || | ||   ||   || ||   | | | | ||||||  |||
Db     221888 GTTGGTTGGTTGGTTGACTGGTTGGTTGGTTGGTTGGTTGGATGGTTGGCTGGCTGGCTG 221829

Qy        351 GCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTT 410
              ||||   ||   || ||   | ||| | | | || ||   ||| |||| |   || | |
Db     221828 GCTGGCTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGACTGACTGGTTG 221769

Qy        411 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 470
              | | | |||    || | | | | | |||   | | | | | | |||||  ||| |||
Db     221768 GTTGGTTGGTTGACTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGGCTGGCTG 221709

Qy        471 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 530
              ||| |||||  || ||| ||| | | | | |||  ||  || | | | |   || ||| ||
Db     221708 GCTGGCTGG-TTGGTTGGTTGGTTGGTTGGCTGGCTGGTTGGTTGGTTGGTT 221650

Qy        531 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTAC 590
              | | |   | ||| ||| | |  |   |   ||||| ||| ||| |  |  ||| |
Db     221649 GGTTGGCTGACTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGCTGACTGGCT 221590

Qy        591 GCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 650
              |  ||  || | | ||||  |   || || | || || | |   || || | |||||
Db     221589 GGTTGGTTGGTTGGTTGACTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCT 221530

Qy        651 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTA 710
              |   |||||   |||||   |   || || | || || | |   ||||    | |||
Db     221529 GGCTGGCTGGCTGGCTGGCTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGACTGACT 221470

Qy        711 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 770
              |   || || | || ||   |   ||||    || || | |   |||||   |||||
Db     221469 GGTTGGTTGGTTGGTTGGCTGGCTGGCTGGCTGGTTGGTTGGTTGGCTGGCTGGCTGACT 221410

Qy        771 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTT 830
              |   |||||   |||||   |   | ||| | || || | |   ||||| |||   | ||
Db     221409 GGTTGGCTGGCTGGCTGGCTGGCTGACTGGTTGGTTGGTTGGCTGGCTGGCTGACTGGTT 221350

Qy        831 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTG 875
              | ||| |||  || | |   ||    |||  || | |     |||||
```

```
Db     221349 GGTTGGTTGGCTGGTTGGTTGGTTCGTTGGTTGGTTGGCTGGCTG 221305


RESULT 51
AC113967/c
LOCUS       AC113967               246968 bp    DNA     linear   HTG 21-OCT-2002
DEFINITION  Mus musculus clone RP23-59I22, WORKING DRAFT SEQUENCE, 3 ordered
            pieces.
ACCESSION   AC113967
VERSION     AC113967.4  GI:24182080
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 246968)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-59I22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 246968)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 246968)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 21, 2002 this sequence version replaced gi:21326293.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
```

```
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
          -------------- Project Information
            Center project name: L23281
            Center clone name: 59_I_22
          -------------- Summary Statistics
            Sequencing vector: Plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 246357 bases at least Q40
            Consensus quality: 246703 bases at least Q30
            Consensus quality: 246763 bases at least Q20
            Insert size: 242000; agarose-fp
            Insert size: 246768; sum-of-contigs
            Quality coverage: 11.2 in Q20 bases; agarose-fp
            Quality coverage: 11.0 in Q20 bases; sum-of-contigs
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 3 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *       1    31803: contig of 31803 bp in length
          *   31804    31903: gap of 100 bp
          *   31904    80596: contig of 48693 bp in length
          *   80597    80696: gap of 100 bp
          *   80697   246968: contig of 166272 bp in length.
FEATURES             Location/Qualifiers
     source          1. .246968
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-59I22"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .31803
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    31904. .80596
                     /note="assembly_fragment"
     misc_feature    80697. .246968
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    73708 a  46388 c  45788 g  80884 t    200 others
ORIGIN

  Query Match             16.6%;  Score 166.2;  DB 2;  Length 246968;
  Best Local Similarity   54.4%;  Pred. No. 9.4e-18;
  Matches  356;  Conservative    0;  Mismatches  298;  Indels    1;  Gaps    1;

Qy        194 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 253
              | | ||| |    | | ||  |     ||||||  |||||  |   ||||  | |||
Db     147975 TCCCTGCCTCCCTGCCTCCCTGCCTCCCTGCCTCCCTGCCTGTCTGCCTCCCTGCCTGTC 147916

Qy        254 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 313
              | | ||  |||||  |||  |   |||||  |||      | | ||  |||| | |||
Db     147915 TGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTCCCTGCC 147856

Qy        314 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 373
                | | |    |||||  ||   ||   |||||  ||          | |   | ||||
Db     147855 TGTCTGTCTGCCTCCCTGCCTGTCTGCCTCCCTGCCTGTCTGCCTCCCTGCCTGTCTGCC 147796

Qy        374 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 433
              | |||  ||  | |||   ||  | |||| ||  |||  |   |||||  ||| ||
Db     147795 TCCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTC 147736

Qy        434 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 493
              || | |||| ||  |||  |   |||||  ||| |||  ||  |||||| ||| |||
Db     147735 TGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCC 147676

Qy        494 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGG 553
              ||  |||||| ||| |||  ||  |||||  ||| |||  ||  |||||| |||
Db     147675 TGCCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGCCTGCCTGCCTGCCTGCCTGCC 147616

Qy        554 GGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTA 613
              | | |   | |||| ||  |||      | | ||  |||||||||   || ||   |
Db     147615 TGCCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGCCTGTCTGCCTGTCTGCC 147556

Qy        614 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 673
              | ||||  | ||||  ||  | |||   | |||   |   | ||||  | ||||  ||
Db     147555 TGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTC 147496

Qy        674 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 733
              | |||   | |||   |   | || |  | ||||  ||  | |||   | ||||  ||
Db     147495 TGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGT-CGTCTGTCTGCCTGCCTGTC 147437

Qy        734 GGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 793
              | ||    | |||   |   | ||||  | ||||  ||  | |||   | |||   |
Db     147436 TGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCC 147377

Qy        794 GGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGT 848
              | ||||  | ||||  ||  | ||| ||||  |  ||| ||  ||  |||  ||
Db     147376 TGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGCCTGTCTGTCTGCCTGT 147322


RESULT 52
AC103472
LOCUS       AC103472               234474 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-208G21, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC103472
VERSION     AC103472.7  GI:30520476
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 234474)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
```

```
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 234474)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 234474)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:25092334.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GHRX
                Center clone name: CH230-208G21
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 223543 bases at least Q40
                Consensus quality: 225303 bases at least Q30
                Consensus quality: 226470 bases at least Q20
                Estimated insert size: 235136; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   233348: contig of 233348 bp in length
            *  233349   233448: gap of unknown length
            *  233449   234474: contig of 1026 bp in length.
FEATURES             Location/Qualifiers
     source          1. .234474
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-208G21"
     misc_feature    1. .2197
                     /note="wgs_contig"
BASE COUNT    62773 a  47967 c  50362 g  65964 t   7408 others
ORIGIN

  Query Match            16.4%;  Score 164.4;  DB 2;  Length 234474;
  Best Local Similarity  53.0%;  Pred. No. 1.8e-17;
  Matches  351;  Conservative    0;  Mismatches  311;  Indels    0;  Gaps    0;

Qy        215 GTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 274
              ||  |  ||||||| ||   ||  |||| | ||     | | | | | || ||| |||
Db      50013 GTGTGTGTGCCTGCGTGTCTGCGCGCCTACGTGTCAGCGTGTCTGCATGCTTGCATGCCT 50072

Qy        275 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 334
              ||  | |||    |  | | | | |   |||||| ||       |   |   |||||| |
Db      50073 GCGTGTCTGTGCACCTATGTGCCTGCGTGCCTGCCTGTGTGCCTGCGTGCCTGCCTGCGT 50132

Qy        335 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTC 394
              ||  ||  || ||| |||    | | | |   |  |||  || ||| |   | ||||
Db      50133 GCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGT 50192

Qy        395 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 454
              ||  |||||| ||| ||  ||  |||||| ||| |||  ||  || ||| ||| |||
Db      50193 GCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGT 50252

Qy        455 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 514
              ||  |||||| ||| ||  |   | |||| ||   |  ||   || ||| ||| ||
Db      50253 GCCTGCCTGCGTGCCTGTCAGTGCGTCTGCGTGTCAGTGCGTCTGCGTGCCTGCGTGTCT 50312

Qy        515 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 574
              ||  |||||| ||| |||  | | |||| | ||||   | | | |   |  ||| ||| |
Db      50313 GCGTGCCTGCGTGCCTGCGTGTTTGCCTACGTGCTTGCGTGTCTGCGTGTGTGCCTGCGT 50372

Qy        575 GCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAG 634
              ||  | | | |     ||||   ||   || ||   |   | ||||  | | ||  ||
Db      50373 GCCTACGTGTCTACGTGCCTATGTGTCAGCGTGCCTGCGTGCCTGCGTGTCAGCGTGTCA 50432

Qy        635 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 694
              |   |  |  ||| |   | ||||  | ||||  ||  |  |||  |  ||  |
Db      50433 GTGCGTCTGCGTGCCTGCGTGTCTGCGTGCCTGCGTGTCTGTATGCCTGCGTGTCTGCGT 50492

Qy        695 GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 754
              | || |  | ||||  |      ||   |  |||       | || |  | ||||  ||
Db      50493 GCCTGCGTGTCTGCGTGCCTCCGTGTCTGCGTGCCTCCGTGTCTGCGTGCCTGCGTGTCT 50552

Qy        755 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 814
              |  |||  |  ||   ||| | ||||  | ||||  |   |  ||   |  |||  |
Db      50553 GTATGCCTGCGTGTCTGTATGCCTGCGTGTCTGCGTGCCTGCGTGTCTGCGTGCCTGCGT 50612

Qy        815 GGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCT 874
              | |||| |||   | ||  ||| | | ||   ||  |   || ||| ||     | | ||
Db      50613 GTCTGCGTGCCTCCGTGTCTGCGTACCTGTGTGTCTGCGTGCCTGCGTGTCTGCATGTCT 50672

Qy        875 GC 876
              ||
Db      50673 GC 50674


RESULT 53
AC112613/c
LOCUS       AC112613              235837 bp    DNA     linear   HTG 13-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-101I20, WORKING DRAFT SEQUENCE, 6
            unordered pieces.
ACCESSION   AC112613
VERSION     AC112613.5  GI:24941847
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
```

```
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 235837)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 235837)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 235837)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 13, 2002 this sequence version replaced gi:23266005.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GRSW
                Center clone name: CH230-101I20
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 223584 bases at least Q40
                Consensus quality: 225354 bases at least Q30
                Consensus quality: 226739 bases at least Q20
                Estimated insert size: 235335; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   184317: contig of 184317 bp in length
            *   184318   184417: gap of unknown length
            *   184418   186305: contig of 1888 bp in length
            *   186306   186405: gap of unknown length
            *   186406   188834: contig of 2429 bp in length
            *   188835   188934: gap of unknown length
            *   188935   192034: contig of 3100 bp in length
            *   192035   192134: gap of unknown length
            *   192135   214096: contig of 21962 bp in length
            *   214097   214196: gap of unknown length
            *   214197   235837: contig of 21641 bp in length.
FEATURES             Location/Qualifiers
     source          1. .235837
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-101I20"
BASE COUNT    69431 a  48015 c  45986 g  63690 t   8715 others
ORIGIN

  Query Match            16.4%;  Score 164.4;  DB 2;  Length 235837;
  Best Local Similarity   53.0%;  Pred. No. 1.8e-17;
  Matches  351;  Conservative    0;  Mismatches  311;  Indels    0;  Gaps    0;

Qy        215 GTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 274
              ||  |  ||||||| ||   ||  |||| | ||     | | | | | || ||| |||
Db     164965 GTGTGTGTGCCTGCGTGTCTGCGCGCCTACGTGTCAGCGTGTCTGCATGCTTGCATGCCT 164906

Qy        275 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 334
              ||  | |||    |  | | | | |   ||||| ||       |   |   |||||| |
Db     164905 GCGTGTCTGTGCACCTATGTGCCTGCGTGCCTGCCTGTGTGCCTGCGTGCCTGCCTGCGT 164846

Qy        335 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTC 394
              ||  ||  || ||| |||    | | | |   | |||  || |||  |   | ||||
Db     164845 GCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGT 164786

Qy        395 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 454
              ||  |||||| ||| |||  ||  |||||| ||| |||  ||  || ||| ||| |||
```

```
Db     164785 GCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGTGCCTGCCTGCGT 164726

Qy        455 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 514
              ||  |||||| ||| ||   |   | |||| ||   |  ||   || ||| ||| ||
Db     164725 GCCTGCCTGCGTGCCTGTCAGTGCGTCTGCGTGTCAGTGCGTCTGCGTGCCTGCGTGTCT 164666

Qy        515 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 574
              ||  |||||| ||| ||| | | |||| | ||||   | | | |   |  ||| ||| |
Db     164665 GCGTGCCTGCGTGCCTGCGTGTTTGCCTACGTGCTTGCGTGTCTGCGTGTGTGCCTGCGT 164606

Qy        575 GCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAG 634
              ||  | | | |     ||||   ||   || ||   |   | ||||  | | ||  ||
Db     164605 GCCTACGTGTCTACGTGCCTATGTGTCAGCGTGCCTGCGTGCCTGCGTGTCAGCGTGTCA 164546

Qy        635 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 694
              |   |   |  ||| |   | |||| | ||||  ||  |  |||  |  ||   |
Db     164545 GTGCGTCTGCGTGCCTGCGTGTCTGCGTGCCTGCGTGTCTGTATGCCTGCGTGTCTGCGT 164486

Qy        695 GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 754
              | || |  | |||| |       ||   |  |||       | || |  | ||||  ||
Db     164485 GCCTGCGTGTCTGCGTGCCTCCGTGTCTGCGTGCCTCCGTGTCTGCGTGCCTGCGTGTCT 164426

Qy        755 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 814
              |  |||  |  ||   ||| | |||| | |||| |   |  ||   |  |||  |
Db     164425 GTATGCCTGCGTGTCTGTATGCCTGCGTGTCTGCGTGCCTGCGTGTCTGCGTGCCTGCGT 164366

Qy        815 GGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCT 874
              | |||| |||   | ||  ||| | | ||   ||  |   || ||| ||     | | ||
Db     164365 GTCTGCGTGCCTCCGTGTCTGCGTACCTGTGTGTCTGCGTGCCTGCGTGTCTGCATGTCT 164306

Qy        875 GC 876
              ||
Db     164305 GC 164304
```

```
RESULT 54
AL845305
LOCUS       AL845305              200141 bp    DNA     linear   ROD 02-DEC-2002
DEFINITION  Mouse DNA sequence from clone RP23-202K13 on chromosome 2, complete
            sequence.
ACCESSION   AL845305
VERSION     AL845305.5  GI:26185503
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 200141)
  AUTHORS   Smith,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 6, 2002 this sequence version replaced gi:23573260.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to confirm this sequence.  Sequence data
            from the whole genome shotgun alone has only been used where it has
            a phred quality of at least 30.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-202K13 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .200141
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-202K13"
                     /clone_lib="RPCI-23"
BASE COUNT    63702 a  37575 c  38312 g  60552 t
ORIGIN
```

```
  Query Match             16.4%;  Score 164;  DB 10;  Length 200141;
  Best Local Similarity   64.5%;  Pred. No. 2.2e-17;
  Matches  245;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy        300 TAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 359
              || || |   |  | |  | |  |   || |||||||| ||  || ||| ||||     |
Db     185417 TAAGCTTTGCTTTTTAATGAAAGGCTTGCTTGCTTGCTTGCCTGCTTGCCTGCTTGCCTG 185476

Qy        360 ACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 419
              | |   | ||||| || |||| ||  |  |||  ||| |||||||||||||||| ||| |
Db     185477 CCTGCCTGCCTGCTTGCCTGCTTGTCTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTG 185536

Qy        420 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 479
              | |||||||||||| ||| || |||||||||||  ||  |||||| ||| |||  ||  |
Db     185537 CTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCCTG 185596

Qy        480 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 539
              ||||| ||| |||  ||| | |||||||| |||| ||  || ||||||||||| ||| |
Db     185597 CCTGCCTGCCTGCCTGCTTGTCTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTGCTTG 185656

Qy        540 CCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTG 599
              |||||||||      | | |   |  ||||||||||||      |   |   || ||| ||
Db     185657 CCTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTG 185716

Qy        600 ATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 659
              | |||||  | |   |  | || | |||||  |   || ||   || ||   | | || ||
Db     185717 CTTGCTTGCTTGCTTGCTTCCTCGCCTGCTTGCTTGGATGGATGGATGGATGGATGGATG 185776

Qy        660 CTGGGCTGCTAGTAGGGCTG 679
                || ||   | | || ||
Db     185777 GATGGATGGATGGATGGATG 185796
```

```
RESULT 55
AC116951
LOCUS       AC116951              256110 bp    DNA     linear   HTG 01-SEP-2002
DEFINITION  Mus musculus chromosome UNK clone RP23-75O6, WORKING DRAFT
            SEQUENCE, 12 unordered pieces.
ACCESSION   AC116951
VERSION     AC116951.2  GI:22539388
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 256110)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
```

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 256110)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 256110)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Aug 29, 2002 this sequence version replaced gi:19909483.

            ----------------- Genome Center ---------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information -------------------
            Center project name: M_BA0075006
            ------------------------------------------------------
            -------------- Summary Statistics  -------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 252203 bases at least Q40
            Consensus quality: 253256 bases at least Q30
            Consensus quality: 254009 bases at least Q20
            Insert size: 216000; agarose-fp
            Insert size: 261658; sum-of-contigs
            Quality coverage: 13.71 in Q20 bases; agarose-fp
            Quality coverage: 9.96 in Q20 bases; sum-of-contigs
            ------------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1235: contig of 1235 bp in length
            *     1236     1335: gap of unknown length
            *     1336     3639: contig of 2304 bp in length
            *     3640     3739: gap of unknown length
            *     3740     5766: contig of 2027 bp in length
            *     5767     5866: gap of unknown length
            *     5867     8394: contig of 2528 bp in length
            *     8395     8494: gap of unknown length
            *     8495    12007: contig of 3513 bp in length
            *    12008    12107: gap of unknown length
            *    12108    17198: contig of 5091 bp in length
            *    17199    17298: gap of unknown length
            *    17299    24643: contig of 7345 bp in length
            *    24644    24743: gap of unknown length
            *    24744    36792: contig of 12049 bp in length
            *    36793    36892: gap of unknown length
            *    36893    59434: contig of 22542 bp in length
            *    59435    59534: gap of unknown length
            *    59535   104072: contig of 44538 bp in length
            *   104073   104172: gap of unknown length
            *   104173   149611: contig of 45439 bp in length
            *   149612   149711: gap of unknown length
            *   149712   256110: contig of 106399 bp in length.
FEATURES             Location/Qualifiers
     source          1. .256110
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-75O6"
     misc_feature    1. .1235
                     /note="assembly_name:Contig16"
     misc_feature    1336. .3639
                     /note="assembly_name:Contig17"
     misc_feature    3740. .5766
                     /note="assembly_name:Contig18"
     misc_feature    5867. .8394
                     /note="assembly_name:Contig19"
     misc_feature    8495. .12007
                     /note="assembly_name:Contig20"
     misc_feature    12108. .17198
                     /note="assembly_name:Contig21"
     misc_feature    17299. .24643
                     /note="assembly_name:Contig22"
     misc_feature    24744. .36792
                     /note="assembly_name:Contig23"
     misc_feature    36893. .59434
                     /note="assembly_name:Contig24"
     misc_feature    59535. .104072
                     /note="assembly_name:Contig25"
     misc_feature    104173. .149611
                     /note="assembly_name:Contig26"
     misc_feature    149712. .256110
                     /note="assembly_name:Contig27"
BASE COUNT    78391 a  49329 c  51220 g  76018 t   1152 others
ORIGIN

  Query Match             16.4%;  Score 164;  DB 2;  Length 256110;
  Best Local Similarity   64.5%;  Pred. No. 2.1e-17;
  Matches  245;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy        300 TAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 359
              || || |   |  | |  | |  |   || |||||||| ||  || ||| ||||     |
Db     219051 TAAGCTTTGCTTTTTAATGAAAGGCTTGCTTGCTTGCTTGCCTGCTTGCCTGCTTGCCTG 219110

Qy        360 ACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 419
               | |   | ||||| || |||| ||  |  |||  ||| ||||||||||||||| ||| |
Db     219111 CCTGCCTGCCTGCTTGCCTGCTTGTCTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTG 219170

Qy        420 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 479
              | |||||||||||| ||| || ||||||||||  ||  |||||| ||| ||| ||  |
Db     219171 CTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCTG 219230

Qy        480 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 539
              ||||| ||| |||  ||| | |||||||| |||| ||  || ||||||||||| ||| |
Db     219231 CCTGCCTGCCTGCCTGCTTGTCTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTGCTTG 219290

Qy        540 CCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTG 599
              |||||||||       | | |   | ||||||||||||      |   |   || ||| ||
Db     219291 CCTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTG 219350

Qy        600 ATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTG 659
              | ||||| | |   |  | || | ||||| |   || ||   || ||   | | || ||
Db     219351 CTTGCTTGCTTGCTTGCTTCCTCGCCTGCTTGCTTGGATGGATGGATGGATGGATGGATG 219410

Qy        660 CTGGGCTGCTAGTAGGGCTG 679
                 ||  ||   | | || ||
Db     219411 GATGGATGGATGGATGGATG 219430


RESULT 56
AC141966
LOCUS       AC141966              197153 bp    DNA     linear   HTG 21-MAR-2003
DEFINITION  Rattus norvegicus clone CH230-372P16, *** SEQUENCING IN PROGRESS
            ***, 73 unordered pieces.
ACCESSION   AC141966
VERSION     AC141966.1  GI:29135423
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
```

REFERENCE 1 (bases 1 to 197153)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 197153)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT ---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: KDFC
Center clone name: CH230-372P16
---------- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 165147 bases at least Q40
Consensus quality: 173510 bases at least Q30
Consensus quality: 180635 bases at least Q20
Estimated insert size: 169130; sum-of-contigs estimation
Quality coverage: 2x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 73 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1049: contig of 1049 bp in length
* 1050 1149: gap of unknown length
* 1150 2432: contig of 1283 bp in length
* 2433 2532: gap of unknown length
* 2533 3890: contig of 1358 bp in length
* 3891 3990: gap of unknown length
* 3991 5466: contig of 1476 bp in length
* 5467 5566: gap of unknown length
* 5567 7037: contig of 1471 bp in length
* 7038 7137: gap of unknown length
* 7138 8152: contig of 1015 bp in length
* 8153 8252: gap of unknown length
* 8253 9570: contig of 1318 bp in length
* 9571 9670: gap of unknown length
* 9671 11147: contig of 1477 bp in length
* 11148 11247: gap of unknown length
* 11248 12667: contig of 1420 bp in length
* 12668 12767: gap of unknown length
* 12768 13786: contig of 1019 bp in length
* 13787 13886: gap of unknown length
* 13887 14962: contig of 1076 bp in length
* 14963 15062: gap of unknown length
* 15063 16631: contig of 1569 bp in length
* 16632 16731: gap of unknown length
* 16732 17738: contig of 1007 bp in length
* 17739 17838: gap of unknown length
* 17839 19324: contig of 1486 bp in length
* 19325 19424: gap of unknown length
* 19425 21063: contig of 1639 bp in length
* 21064 21163: gap of unknown length
* 21164 22382: contig of 1219 bp in length
* 22383 22482: gap of unknown length
* 22483 23824: contig of 1342 bp in length
* 23825 23924: gap of unknown length
* 23925 25794: contig of 1870 bp in length
* 25795 25894: gap of unknown length
* 25895 27228: contig of 1334 bp in length
* 27229 27328: gap of unknown length
* 27329 28469: contig of 1141 bp in length
* 28470 28569: gap of unknown length
* 28570 29748: contig of 1179 bp in length
* 29749 29848: gap of unknown length
* 29849 31625: contig of 1777 bp in length
* 31626 31725: gap of unknown length
* 31726 32950: contig of 1225 bp in length
* 32951 33050: gap of unknown length
* 33051 34330: contig of 1280 bp in length
* 34331 34430: gap of unknown length
* 34431 36256: contig of 1826 bp in length
* 36257 36356: gap of unknown length
* 36357 37893: contig of 1537 bp in length
* 37894 37993: gap of unknown length
* 37994 39984: contig of 1991 bp in length
* 39985 40084: gap of unknown length
* 40085 41384: contig of 1300 bp in length
* 41385 41484: gap of unknown length
* 41485 43490: contig of 2006 bp in length
* 43491 43590: gap of unknown length
* 43591 45354: contig of 1764 bp in length
* 45355 45454: gap of unknown length
* 45455 46885: contig of 1431 bp in length

```
            *    46886    46985: gap of unknown length
            *    46986    48685: contig of 1700 bp in length
            *    48686    48785: gap of unknown length
            *    48786    50073: contig of 1288 bp in length
            *    50074    50173: gap of unknown length
            *    50174    51173: contig of 1000 bp in length
            *    51174    51273: gap of unknown length
            *    51274    52907: contig of 1634 bp in length
            *    52908    53007: gap of unknown length
            *    53008    55111: contig of 2104 bp in length
            *    55112    55211: gap of unknown length
            *    55212    57269: contig of 2058 bp in length
            *    57270    57369: gap of unknown length
            *    57370    58856: contig of 1487 bp in length
            *    58857    58956: gap of unknown length
            *    58957    61331: contig of 2375 bp in length
            *    61332    61431: gap of unknown length
            *    61432    63670: contig of 2239 bp in length
            *    63671    63770: gap of unknown length
            *    63771    66965: contig of 3195 bp in length
            *    66966    67065: gap of unknown length
            *    67066    68564: contig of 1499 bp in length
            *    68565    68664: gap of unknown length
            *    68665    70885: contig of 2221 bp in length
            *    70886    70985: gap of unknown length
            *    70986    73327: contig of 2342 bp in length
            *    73328    73427: gap of unknown length
            *    73428    77032: contig of 3605 bp in length
            *    77033    77132: gap of unknown length
            *    77133    79086: contig of 1954 bp in length
            *    79087    79186: gap of unknown length
            *    79187    80549: contig of 1363 bp in length
            *    80550    80649: gap of unknown length
            *    80650    82568: contig of 1919 bp in length
            *    82569    82668: gap of unknown length
            *    82669    85166: contig of 2498 bp in length
            *    85167    85266: gap of unknown length
            *    85267    87990: contig of 2724 bp in length
            *    87991    88090: gap of unknown length
            *    88091    90803: contig of 2713 bp in length
            *    90804    90903: gap of unknown length
            *    90904    94173: contig of 3270 bp in length
            *    94174    94273: gap of unknown length

  Query Match             16.2%;  Score 162.6;  DB 2;  Length 197153;
  Best Local Similarity   65.3%;  Pred. No. 3.7e-17;
  Matches  254;  Conservative    0;  Mismatches  134;  Indels    1;  Gaps    1;

Qy        189 GCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGC 248
              |||  |   |||||||  |    |  |   | ||||||| |||| ||| || ||| |||
Db     111170 GCGCTTTGCTGCTTGTTTGCTTGCTTGCCTGCCTGCCTGCTTGCTTGCTTGCTTGCCTGC 111229

Qy        249 TGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGC 308
                | | |   | | || |||| | | || ||| |||      |   |   |  |||
Db     111230 CTGCCTGCCTGCTTGTCGGCCTGCTTGTTTGCTTGCCTGCCTGCTTGCTTGCTTGTTTGC 111289

Qy        309 TTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGG 368
              |||||     | | |   |||||||||| ||  ||||||||||     |   ||  |
Db     111290 TTGCTTGCCTGCCTGCCTGCCTGCTTGCTTGCCTGCCTGCTTGCTTGTTTGCTTGTCTGC 111349

Qy        369 CTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 428
              ||||  |||||||  ||  |  ||   ||  ||||||||||||||  ||  || |||||||
Db     111350 CTGCCTGCTTGCTTGTTTGCTTGTCTGCCTGCCTGCTTGCTTGCCTGCCTGCTTGCTTGC 111409

Qy        429 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 488
              ||||  ||  ||||||||||||||  ||  ||||||||||||||  ||  |||||| |||
Db     111410 TTGCCTGCCTGCCTGCTTGCTTGCCTGCCTGCCTGCTTGCTTGCCTGCCTGCCTGCCTGC 111469

Qy        489 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 548
              |||  ||| || |||||| ||||| ||| |  ||||||||||| ||| || || ||||
Db     111470 CTGCCTGCTTGCTTGCTTGTTTGCTTGCTTGTTTGCTTGCTTGCT-GCTTGCTTGTTTGC 111528

Qy        549 TGAGGGGACAGTAGGGCTGCTTGCTTGCT 577
```

```
              |   |   |   | ||||||||||||
Db     111529 TTGCTTGTTTGCCTGCCTGCTTGCTTGCT 111557


RESULT 57
AC094273/c
LOCUS       AC094273              228645 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-3G15, *** SEQUENCING IN PROGRESS ***,
            2 unordered pieces.
ACCESSION   AC094273
VERSION     AC094273.6  GI:30467504
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 228645)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 228645)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
```

```
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
REFERENCE  3  (bases 1 to 228645)
  AUTHORS  Rat Genome Sequencing Consortium.
  TITLE    Direct Submission
  JOURNAL  Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
COMMENT    On May 9, 2003 this sequence version replaced gi:24955409.
           The sequence in this assembly is a combination of BAC based reads
           and whole genome shotgun sequencing reads assembled using Atlas
           (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
           in the feature table below represents a scaffold in the Atlas
           assembly (a 'contig-scaffold'). Within each contig-scaffold,
           individual sequence contigs are ordered and oriented, and separated
           by sized gaps filled with Ns to the estimated size. The sequence
           may extend beyond the ends of the clone and there may be sequence
           contigs within a contig-scaffold that consist entirely of whole
           genome shotgun sequence reads. Both end sequences and whole genome
           shotgun sequence only contigs will be indicated in the feature
           table.
           -------------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
           -------------- Project Information
               Center project name: GAHS
               Center clone name: CH230-3G15
           -------------- Summary Statistics
               Assembly program: Atlas;
               Consensus quality: 217014 bases at least Q40
               Consensus quality: 220400 bases at least Q30
               Consensus quality: 223003 bases at least Q20
               Estimated insert size: 227771; sum-of-contigs estimation
               Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
           --------------
           * NOTE: Estimated insert size may differ from sequence length
           *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 2 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           *        1   226692: contig of 226692 bp in length
           *   226693   226792: gap of unknown length
           *   226793   228645: contig of 1853 bp in length.
FEATURES             Location/Qualifiers
     source          1. .228645
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-3G15"
     misc_feature    1. .2101
                     /note="wgs_contig"
BASE COUNT    54911 a  52298 c  54612 g  62695 t   4129 others
ORIGIN

  Query Match             16.2%;  Score 162.6;  DB 2;  Length 228645;
  Best Local Similarity   65.3%;  Pred. No. 3.6e-17;
  Matches  254;  Conservative    0;  Mismatches  134;  Indels    1;  Gaps    1;

Qy        189 GCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGC 248
              |||  |   |||||||  |    |  |   | ||||||| |||| ||| || ||| |||
Db       8080 GCGCTTTGCTGCTTGTTTGCTTGCTTGCCTGCCTGCCTGCTTGCTTGCTTGCTTGCCTGC 8021

Qy        249 TGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGC 308
                | | |   | | || |||| | | || ||| |||      |   |   | |||
Db       8020 CTGCCTGCCTGCTTGTCGGCCTGCTTGTTTGCTTGCCTGCCTGCTTGCTTGCTTGTTTGC 7961

Qy        309 TTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGG 368
              |||||     | | |   |||||||||| ||  ||||||||||     |   ||  |
Db       7960 TTGCTTGCCTGCCTGCCTGCCTGCTTGCTTGCCTGCCTGCTTGCTTGTTTGCTTGTCTGC 7901

Qy        369 CTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 428
              ||||  ||||||| ||  |  ||   ||  |||||||||||||  ||  || |||||||
Db       7900 CTGCCTGCTTGCTTGTTTGCTTGTCTGCCTGCCTGCTTGCTTGCCTGCCTGCTTGCTTGC 7841

Qy        429 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 488
              ||||  ||  |||||||||||||  ||  |||||||||||||  ||  |||||| |||
Db       7840 TTGCCTGCCTGCCTGCTTGCTTGCCTGCCTGCCTGCTTGCTTGCCTGCCTGCCTGCCTGC 7781

Qy        489 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 548
               |||  ||| || |||||| ||||| ||| |  ||||||||||| ||| || || ||||
Db       7780 CTGCCTGCTTGCTTGCTTGTTTGCTTGCTTGTTTGCTTGCTTGCT-GCTTGCTTGTTTGC 7722

Qy        549 TGAGGGGACAGTAGGGCTGCTTGCTTGCT 577
              |     |   |   | ||||||||||||
Db       7721 TTGCTTGTTTGCCTGCCTGCTTGCTTGCT 7693
```

```
RESULT 58
AC095211
LOCUS       AC095211               252801 bp    DNA     linear   HTG 03-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-9E3, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC095211
VERSION     AC095211.4  GI:22773101
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 252801)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
```

```
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 252801)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 252801)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 10, 2002 this sequence version replaced gi:21722802.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequening reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/).  As a result, the
            sequence may extend beyond the ends of the clone and there may be
            contigs that consist entirely of whole genome shotgun sequence
            reads.  Both end sequences and whole genome shotgun sequence only
            contigs will be indicated in the feature table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GDHZ
                Center clone name: CH230-9E3
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 222989 bases at least Q40
                Consensus quality: 225848 bases at least Q30
                Consensus quality: 227780 bases at least Q20
                Estimated insert size: 241532; sum-of-contigs estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   252801: contig of 252801 bp in length.
FEATURES            Location/Qualifiers
     source         1. .252801
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-9E3"
     misc_feature   1. .1620
                    /note="wgs_contig"
     misc_feature   251753. .252801
                    /note="wgs_contig"
BASE COUNT    63531 a  49651 c  49432 g  66454 t  23733 others
ORIGIN
```

```
  Query Match           16.1%;  Score 161.4;  DB 2;  Length 252801;
  Best Local Similarity   54.8%;  Pred. No. 5.4e-17;
  Matches  361;  Conservative    0;  Mismatches  296;  Indels    2;  Gaps    2;

Qy        191 GTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTG 250
              ||  | | ||| ||  |    | |||  |   |   ||  ||| | |||   | | | ||
Db     185211 GTTGGTGGGCTGGTTGGTTGGCTGGTTGGTTGGTTGGCTGGCTGGTTGGTTGGTTGGTTG 185270

Qy        251 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 310
                ||   || ||   | | | | | |||   ||| | ||   || |||| ||  || ||
Db     185271 GTTGGTTGGTTGGTTGGTTGGTTGGTTGGTGGGCTGGTTGGTTGGTCAGTCGG-TTGGTT 185329

Qy        311 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 370
              | |  | || | |   |  || ||| | | ||| ||| ||| |  |  |   |   || |
Db     185330 GGTTGGTGGGCTGGTTGGTTGGTTGGTTGGTGGGCTGGTTGGTTGGTTGGTTGGTTGGTT 185389

Qy        371 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 430
              | | | |||   |   || || ||  | |  || ||| |||   | | |  ||  || ||
Db     185390 GGTTGGTTGGCTGGTTGGTTGGTCAGTCGGTTGGTTGGTTGGCTGGTTGGTTGACTGGTT 185449

Qy        431 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              | | |   | ||||||| |||   | | || || | |  || | | | | ||| ||| ||
Db     185450 GGTTGGCTGGTTGCTTGGTTGGCTGGTTGCTTGGTAGGCTGATTGGTTGGCTGGTTGGTT 185509

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              | | | ||| ||| ||| |||   | | || || ||| ||||| | ||| ||| |||||
Db     185510 GGTTGGTGGGCTGATTGGTTGGCTGGTTGCTTGGTTGGTTGCTTGGTGGGCTGGTTGCTT 185569

Qy        551 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 610
              | || |     || ||  || ||||| |  | | |   | |||   | |||   | |
Db     185570 GGTGGGCTTGTTGGTTGGCTGGTTGCTTGGTTGGCTGGTGGGCTGTTGGTTGGTGGGCTG 185629

Qy        611 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 670
              | ||   ||||   ||| || ||  || | || || | |   || || | ||||| |
Db     185630 ATTGGTTGGCTGGTTGGCTGGTTGGT-TGGTTGGTTGGTGGGCTGGTTGTTTGGCTGGTT 185688

Qy        671 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 730
              |   ||  |    | || | |  ||||  | || || | |   || || |   |||| |
Db     185689 GGTTGGTGGACTAGTTGGTTGGTGGGCTGGTTGGTTGGTTGGCTGGTTGGTTAGCTGGTT 185748

Qy        731 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 790
              |   ||||  | || || | |   || || | ||||| | |   ||||| | || || |
Db     185749 GGTTGGCTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGTT 185808

Qy        791 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTAGTG 849
              |   |  |   || || | |   || || ||| |   ||| | |   || ||| | .||
Db     185809 GGCTAGTGGGCTGGTTGGTTGGTTGGTTGGCTGGTTAGTTGATGGGCTGGTTGGTTTTG 185867
```

```
RESULT 59
AC115149
LOCUS       AC115149               242482 bp    DNA     linear   HTG 13-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-95C4, WORKING DRAFT SEQUENCE.
ACCESSION   AC115149
VERSION     AC115149.4  GI:24941931
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 242482)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
```

Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 242482)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (15-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 242482)

AUTHORS Rat Genome Sequencing Consortium.

TITLE Direct Submission

JOURNAL Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Nov 13, 2002 this sequence version replaced gi:23665230.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

---------- Project Information
Center project name: GOPR
Center clone name: CH230-95C4

---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 222780 bases at least Q40
Consensus quality: 224645 bases at least Q30
Consensus quality: 225630 bases at least Q20
Estimated insert size: 230869; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

---------- * NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 242482: contig of 242482 bp in length.

FEATURES Location/Qualifiers

source 1. .242482
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-95C4"

misc_feature 1. .1190
/note="wgs_end_extension
clone_end:Sp6"

misc_feature 1692. .3431
/note="wgs_end_extension
clone_end:Sp6"

misc_feature 4820. .5678
/note="clone_boundary
clone_end:Sp6
site:
end_sequence:BH357978"

BASE COUNT 56337 a 54961 c 55419 g 59910 t 15855 others

ORIGIN

Query Match 16.1%; Score 161; DB 2; Length 242482;
Best Local Similarity 52.9%; Pred. No. 6.3e-17;
Matches 370; Conservative 0; Mismatches 325; Indels 4; Gaps 1;

```
Qy       167 GGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCT 226
            | | | || |    || | | |||||| | ||  |   | ||| || | |
Db    217882 GTGTGGTGGCGTGGTGGTGTGGTAGTGTGCTGGTGTGGTGGTGTGCTGGTGTGGTGGTGT 217941

Qy       227 GCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTT 286
            ||  ||  | |||  |||| ||   | ||   |  ||  || | ||  | |||  ||||
Db    217942 GCTGGCGTGGTGGTGTGCTGGCGTGGTGGTGTGCTGGTGTGGTGGCATGGTGGTGTGCTG 218001

Qy       287 GCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCT 346
            ||   | ||   |  ||| || | ||  |   |  | ||  |||| ||  | |||  |
Db    218002 GCGTGGTGGTGTGCTGGCGTGGTGGCGTAAAAAAGTGGTGGTGTGCTGGCGTGGTGGTGT 218061

Qy       347 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 406
            ||| |    | ||   |  ||  || |||| || | ||  |   ||  | |||| || |
Db    218062 GCTGGTGTGGTGGTGTGCTGGTGTGGTGGCATGGTGGTGTGCTGGCGTGGTGGCGTGGTG 218121

Qy       407 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 466
            || || | ||  ||  || || | ||  | |||| || | || || | ||  |   || |
Db    218122 GCGTGGTGGCGTGCTGGCGTGGTGGCGTGGTGGCGTGGTGGCGTGGTGGCGTGGTGGCGT 218181

Qy       467 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 526
            | | ||  |||||| || | || |||| ||  |   |  |||| ||  | |||  ||||
Db    218182 GGTGGCGTGCTGGCGTGGTGGCGTGCTGGCGTGGTGGTGTGCTGGCGTGGTGGTGTGCTG 218241

Qy       527 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACG 586
            || |    | |||  |||| ||   | ||   |  ||  || | || || |   | | |
Db    218242 GCGT----GGTGGTGTGCTGGCGTGGTGGCGTGCTGGCGTGGTGGCGTGGTGGCGTGGTG 218297

Qy       587 GTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 646
```

```
Db      218298 GCGTGCTGGCGTGGTGGTGTGCTGGCGTGCTGGCGTGGTGGCGTGCTGGCGTGGTGGTGT 218357

Qy         647 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCT 706

Db      218358 GCTGGCGTGGTGGTGTGGTGGCGTGGTGGTGTGCTGGTGTGGTGGTGTGCTGGTGTGGTG 218417

Qy         707 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCT 766

Db      218418 GCATGGTGGTGTGCTGGTGTGGTGGCATGGTGGCGTGGTGGCGTGGTGGCGTGCTGGCGT 218477

Qy         767 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTG 826

Db      218478 GGTGGCGTGCTGGCGTGGTGGCGTGCTGGCGTGGTGGCGTGCTGGCGTGGTGGTGTGCTG 218537

Qy         827 GCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCT 865

Db      218538 GCGTGGTGGTGTGCTGGCGTGGTGGTGTGCTGGCGTGGT 218576


RESULT 60
AC094136/c
LOCUS       AC094136               257675 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-3D20, *** SEQUENCING IN PROGRESS ***,
            3 unordered pieces.
ACCESSION   AC094136
VERSION     AC094136.6  GI:30467923
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 257675)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 257675)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 257675)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 9, 2003 this sequence version replaced gi:24819678.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GABE
                Center clone name: CH230-3D20
            -------------- Summary Statistics
                Assembly program: Atlas;
                Consensus quality: 233178 bases at least Q40
                Consensus quality: 236706 bases at least Q30
                Consensus quality: 239237 bases at least Q20
                Estimated insert size: 245474; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   254381: contig of 254381 bp in length
            *  254382   254481: gap of unknown length
            *  254482   255929: contig of 1448 bp in length
            *  255930   256029: gap of unknown length
            *  256030   257675: contig of 1646 bp in length.
FEATURES             Location/Qualifiers
     source          1. .257675
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
```

```
                     /clone="CH230-3D20"
     misc_feature    1. .1653
                     /note="wgs_contig"
     misc_feature    6450. .8473
                     /note="wgs_contig"
     misc_feature    223108. .225095
                     /note="wgs_contig"
     misc_feature    complement(246438. .247075)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH308583"
     misc_feature    252628. .254381
                     /note="wgs_end_extension
                     clone_end:T7"
BASE COUNT    68663 a  49595 c  49881 g  72963 t  16573 others
ORIGIN

  Query Match           16.1%;  Score 161;  DB 2;  Length 257675;
  Best Local Similarity   58.8%;  Pred. No. 6.2e-17;
  Matches  278;  Conservative    0;  Mismatches  195;  Indels    0;  Gaps    0;

Qy        191 GTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTG 250
              |||||  ||| || | | |  || |   |  ||||||| |||  ||  ||||||| |||
Db      88726 GTGTGACTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCT 88667

Qy        251 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 310
                | | | |   |||||| |||  ||  |||||| |||    | | | |   |||||| |
Db      88666 GCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGT 88607

Qy        311 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 370
              ||    | | | |   |||||| |||  ||  |||||| |||    | | | |   | ||
Db      88606 GCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCT 88547

Qy        371 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 430
              ||  || |||  |   |  |||  ||  |||||| ||| ||| ||  || ||| ||| |
Db      88546 GCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGT 88487

Qy        431 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              ||  ||  |||||| ||| |||  || .|| ||| ||| |||  ||  |||||| ||| |
Db      88486 GCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCT 88427

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              ||  |.|. || ||| ||| |||  ||  |||||| ||| |||  ||  || ||| |||
Db      88426 GCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGT 88367

Qy        551 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 610
                  |   |   | ||| ||| |||    | | | |   |||||| ||   || ||
Db      88366 GCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCT 88307

Qy        611 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 663
              |   | ||||  | ||||  |   |  |||  |  ||| | |    || | ||
Db      88306 GCGTGCCTGCGTGCCTGCGTGCCTGCGTGCCTGCGTGCGACTGCCTCTACAGG 88254


RESULT 61
AC118905
LOCUS       AC118905               205515 bp    DNA     linear   HTG 12-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-255D17, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC118905
VERSION     AC118905.7  GI:23908079
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 205515)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 205515)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 205515)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 12, 2002 this sequence version replaced gi:21902607.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ---------- Genome Center
```

```
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
            Center project name: GWJB
            Center clone name: CH230-255D17
          -------------- Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 184541 bases at least Q40
            Consensus quality: 186580 bases at least Q30
            Consensus quality: 188187 bases at least Q20
            Estimated insert size: 190219; sum-of-contigs estimation
            Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 4 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1    22234: contig of 22234 bp in length
          *    22235    22334: gap of unknown length
          *    22335   154198: contig of 131864 bp in length
          *   154199   154298: gap of unknown length
          *   154299   179162: contig of 24864 bp in length
          *   179163   179262: gap of unknown length
          *   179263   205515: contig of 26253 bp in length.
FEATURES             Location/Qualifiers
     source          1. .205515
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-255D17"
     misc_feature    1. .1336
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    complement(938. .1516)
                     /note="clone_boundary
                     clone_end:T7
                     site:MboI
                     end_sequence:RXACK21TJ"
     misc_feature    51900. .52907
                     /note="wgs_contig"
BASE COUNT    56529 a  39946 c  39535 g  53910 t  15595 others
ORIGIN

  Query Match            16.0%;  Score 160.6;  DB 2;  Length 205515;
  Best Local Similarity  64.3%;  Pred. No. 7.6e-17;
  Matches  241;  Conservative    0;  Mismatches  134;  Indels    0;  Gaps    0;

Qy        191 GTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTG 250
              || ||| ||| ||   |    |  ||  | |||  |||||| | ||  || ||| |  ||
Db     121932 GTTTGCCTGCCTGCCTGCTTGCCTGTTTGCCTGTTTGCCTGTTTGCCTGCTTGCGTCTTG 121991

Qy        251 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 310
                |       |       |  || | | ||||||||||      | | |   || |||||
Db     121992 CTTGCTTGCCTGCTTGCTTGCCTTGTTTGCCTGCTTGCCTGCTTGCCTGCTTGCTTGCTT 122051

Qy        311 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 370
              |    |   |  ||||||||||||||| ||||||||||     | | |   |  |
Db     122052 GTCTGCTTGCTCGCTTGCCTGCTTGCTCGCTTGCCTGCTTGCTTGCCTGCCTGCCTGCTT 122111

Qy        371 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 430
              ||| ||||||  |   |  |||| ||| || ||| |||||||  ||| ||||||||||
Db     122112 GCTCGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCT 122171

Qy        431 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              ||| ||| || ||| |||||||  ||  |||||||||||| ||||||  ||  || |||||

Db     122172 GCTTGCTCGCTTGCCTGCTTGCCTGCCTGCCTGCTTGCTCGCTCGCCTGCTCGCCTGCTT 122231

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              ||| ||| |||||||||| |||| ||  ||  ||||||||  ||||||  || |||| |||
Db     122232 GCTTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCTTGCTCACTCGCTTGCTTGCTCGCTC 122291

Qy        551 AGGGGACAGTAGGGC 565
              |       || |||||
Db     122292 ATTTTGAAGCAGGGC 122306


RESULT 62
AC109709
LOCUS       AC109709              272652 bp    DNA     linear   HTG 09-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-228K2, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
ACCESSION   AC109709
VERSION     AC109709.4  GI:23195460
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 272652)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
```

```
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 272652)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 272652)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 19, 2002 this sequence version replaced gi:21738042.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ---------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
                Center project name: GQEA
                Center clone name: CH230-228K2
            ---------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 233404 bases at least Q40
                Consensus quality: 236957 bases at least Q30
                Consensus quality: 239428 bases at least Q20
                Estimated insert size: 252351; sum-of-contigs estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   170363: contig of 170363 bp in length
            *  170364   170463: gap of unknown length
            *  170464   271425: contig of 100962 bp in length
            *  271426   271525: gap of unknown length
            *  271526   272652: contig of 1127 bp in length.
FEATURES             Location/Qualifiers
     source          1. .272652
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-228K2"
     misc_feature    complement(95812. .96229)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:RWBNV61TJ"
     misc_feature    152380. .153698
                     /note="wgs_contig"
     misc_feature    complement(168487. .169398)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:RWBNV61TV"
     misc_feature    170464. .172041
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    266872. .268448
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    268995. .271425
                     /note="wgs_end_extension
                     clone_end:Sp6"
BASE COUNT    71259 a  49098 c  50266 g  71245 t  30784 others
ORIGIN

  Query Match             16.0%;  Score 160.6;  DB 2;  Length 272652;
  Best Local Similarity   64.3%;  Pred. No. 7.1e-17;
  Matches  241;  Conservative    0;  Mismatches  134;  Indels    0;  Gaps    0;

Qy        191 GTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTG 250
              || ||| ||| ||   |    |  ||  | |||  |||||| | ||  || ||| |  ||
Db     140431 GTTTGCCTGCCTGCCTGCTTGCCTGTTTGCCTGTTTGCCTGTTTGCCTGCTTGCGTCTTG 140490

Qy        251 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 310
              |      |     | || | | ||||||||||      | | |   || |||||
Db     140491 CTTGCTTGCCTGCTTGCTTGCCTTGTTTGCCTGCTTGCCTGCTTGCCTGCTTGCTTGCTT 140550

Qy        311 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 370
              |       |   |  |||||||||||||| ||||||||||     | | |   | |
Db     140551 GTCTGCTTGCTCGCTTGCCTGCTTGCTCGCTTGCCTGCTTGCTTGCCTGCCTGCCTGCTT 140610

Qy        371 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 430
              ||| ||||||  |   |  |||| ||| || ||| |||||||  ||| ||||||||| |
Db     140611 GCTCGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCT 140670

Qy        431 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              ||| ||| || ||| |||||||  ||  |||||||||||| ||||||  ||  || |||||
Db     140671 GCTTGCTCGCTTGCCTGCTTGCCTGCCTGCCTGCTTGCTCGCTCGCCTGCTCGCCTGCTT 140730

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              ||| ||| |||||||||| |||| ||  ||  ||||||||  |||||| || |||| |||
Db     140731 GCTTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCTTGCTCACTCGCTTGCTTGCTCGCTC 140790

Qy        551 AGGGGACAGTAGGGC 565
              |      || |||||
Db     140791 ATTTTGAAGCAGGGC 140805


RESULT 63
AC118832
LOCUS       AC118832              206805 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-271G9, *** SEQUENCING IN PROGRESS
            ***, 12 unordered pieces.
ACCESSION   AC118832
VERSION     AC118832.4  GI:25094002
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 206805)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
```

```
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 206805)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 206805)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:23908265.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GVTR
                Center clone name: CH230-271G9
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 172120 bases at least Q40
                Consensus quality: 175720 bases at least Q30
                Consensus quality: 178116 bases at least Q20
                Estimated insert size: 177937; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    48963: contig of 48963 bp in length
            *    48964    49063: gap of unknown length
            *    49064    77623: contig of 28560 bp in length
            *    77624    77723: gap of unknown length
            *    77724    88733: contig of 11010 bp in length
            *    88734    88833: gap of unknown length
            *    88834   169408: contig of 80575 bp in length
            *   169409   169508: gap of unknown length
            *   169509   180885: contig of 11377 bp in length
            *   180886   180985: gap of unknown length
            *   180986   191517: contig of 10532 bp in length
            *   191518   191617: gap of unknown length
            *   191618   193312: contig of 1695 bp in length
            *   193313   193412: gap of unknown length
            *   193413   194460: contig of 1048 bp in length
            *   194461   194560: gap of unknown length
            *   194561   195655: contig of 1095 bp in length
            *   195656   195755: gap of unknown length
            *   195756   198673: contig of 2918 bp in length
            *   198674   198773: gap of unknown length
            *   198774   202107: contig of 3334 bp in length
            *   202108   202207: gap of unknown length
            *   202208   206805: contig of 4598 bp in length.
FEATURES             Location/Qualifiers
     source          1. .206805
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-271G9"
     misc_feature    1. .1288
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    complement(9131. .9885)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BZ218662"
     misc_feature    164162. .165223
                     /note="wgs_contig"
     misc_feature    169509. .171921
                     /note="wgs_contig"
     misc_feature    180986. .182266
                     /note="wgs_contig"
     misc_feature    185783. .187052
                     /note="wgs_contig"
BASE COUNT    53759 a  37440 c  37301 g  51783 t  26522 others
ORIGIN

  Query Match             15.9%;  Score 159.2;  DB 2;  Length 206805;
  Best Local Similarity   54.0%;  Pred. No. 1.3e-16;
  Matches  343;  Conservative    0;  Mismatches  291;  Indels    1;  Gaps    1;

Qy        231 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 290
```

```
Db     154169 GTTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGACTGGCTG 154228

Qy        291 AGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTT 350

Db     154229 GTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGCTGGTTGGTTGGTTG 154288

Qy        351 GCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTT 410

Db     154289 GTTGGTTGGCTGGTTGGTTGGCTGGTTGGCTGGTTGGTTGGTTGGTTGACTGGCTGGCTG 154348

Qy        411 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 470

Db     154349 GTTGGTTGG-TTGGTTGGTTGGTTGGTTGGCTGGTTGGCTGGTTGGTTGGCTGGTTGGCT 154407

Qy        471 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 530

Db     154408 GGCTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGTTGGTT 154467

Qy        531 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTAC 590

Db     154468 GGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGACTGACTGGTT 154527

Qy        591 GCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 650

Db     154528 GGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGCTGGTTGGTTGGTT 154587

Qy        651 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTA 710

Db     154588 GGTTGACTGACTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGCTGGTT 154647

Qy        711 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 770

Db     154648 GGCTGGTTGGTTGGTTGGTTGGCTGGTTGGCTGGTTGGCTGGTTGGCTGTTTTGTTGATT 154707

Qy        771 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTT 830

Db     154708 CGCTGGTTGGTTGGTTGTTTGGTTGGTTGGCTGGTTGGTTGTTTGGTTGGCTAGTTNGTT 154767

Qy        831 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCT 865

Db     154768 GCNTTGTTGGCTCGTTGTTTGGTTGGTAGNCTGGT 154802
```

```
RESULT 64
AC106350/c
LOCUS       AC106350               256486 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-94P16, WORKING DRAFT SEQUENCE, 6
            unordered pieces.
ACCESSION   AC106350
VERSION     AC106350.4  GI:30581308
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 256486)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 256486)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 256486)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:23268877.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GKSS
                Center clone name: CH230-94P16
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 236501 bases at least Q40
                Consensus quality: 239568 bases at least Q30
```

```
            Consensus quality: 241652 bases at least Q20
            Estimated insert size: 246612; sum-of-contigs estimation
            Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 6 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1    11137: contig of 11137 bp in length
          *   11138    11237: gap of unknown length
          *   11238   249465: contig of 238228 bp in length
          *  249466   249565: gap of unknown length
          *  249566   250639: contig of 1074 bp in length
          *  250640   250739: gap of unknown length
          *  250740   251829: contig of 1090 bp in length
          *  251830   251929: gap of unknown length
          *  251930   253165: contig of 1236 bp in length
          *  253166   253265: gap of unknown length
          *  253266   256486: contig of 3221 bp in length.
FEATURES             Location/Qualifiers
     source          1. .256486
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-94P16"
     misc_feature    1. .1811
                     /note="wgs_contig"
     misc_feature    11238. .12909
                     /note="wgs_contig"
     misc_feature    243988. .246089
                     /note="wgs_contig"
     misc_feature    247863. .249465
                     /note="wgs_contig"
BASE COUNT    65807 a  51132 c  53326 g  72818 t  13403 others
ORIGIN

  Query Match           15.8%;  Score 158;  DB 2;  Length 256486;
  Best Local Similarity   52.2%;  Pred. No. 1.9e-16;
  Matches  350;  Conservative    0;  Mismatches  320;  Indels    0;  Gaps    0;

Qy        219 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 278
              || ||  ||  || | | | |  |   || |  |  |   |   |  || ||| | | |
Db      83984 GGTTGGTTGGTTGGTTGGTTGGTTTGATGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 83925

Qy        279 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 338
              |  || ||| |  |  |   | | |  || ||| |  |  |   |   |  || ||| |
Db      83924 GGTTGGTTGGTTTGATGGTTGGATGGTTGGTTGGTTGGATGGTTGGTTGGTTGGTT 83865

Qy        339 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 398
              |   |  || |||    |  |   |   || || | | ||| | |   ||||| | | |
Db      83864 GGATGGATGGTTGGATGGTTGGTTGGTTGGATGGTTGGTTGGTTGGCTGGTTGGTT 83805

Qy        399 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 458
              |  |  ||||| | | | |  ||  || ||| | | | |  || |||  ||   | |
Db      83804 GGTTTGATGCTTGGTTGGTTGGTTGGGTGGTTGGTTGGTTGTTTGGTTGGATGGATGGTT 83745

Qy        459 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 518
              |  || ||| ||| | | | || |||||||||||| ||| || ||||||  |  | | |
Db      83744 GGATGGTTGGTTGGTTGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGGATTGTTGGTT 83685

Qy        519 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTA 578
              |  || ||| ||| | | | |  || ||| |  |  |   | | || || ||| ||| |
Db      83684 GGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGATGGATGGTTGGTTGGTT 83625

Qy        579 AGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 638
              |  |   | | |  ||  || | | ||| | | | || || | || || | |   || |
Db      83624 GGTTGGATGGATGGTTGGATGGTTGGTTGGTTGGATGGTTGGTTGGTTGGCTGGTT 83565

Qy        639 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 698
              | | || |    |   || || | || ||   |   || || | |  || | | | || |
Db      83564 GGTTGGTTTGATGCTTGGTTGGTTGGTTGGGTGGTTGGTTGGTTGCTTGGTTGGATGGAT 83505

Qy        699 CCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCT 758
                | |  || | |   || |||| |  |||| |   |  | || |  |||| |   |  |
Db      83504 GGTTGCATGGTTGGTTGGTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 83445

Qy        759 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 818
              |    | || | |   || || | || || | |   || || | || ||   | | || |
Db      83444 GGATTGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGATGGATGGTT 83385

Qy        819 GCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTG 878
              |  || | |  || |||  || ||| | |   ||  | |||  || |     || || |
Db      83384 GGTTGGTTGGATGGTTGGATGGTTGGTTGGTTGGTTGGTTGGATGGTTGGTTGGTTGGTT 83325

Qy        879 TGCAGCTGGG 888
              |  | | ||
Db      83324 GGATGGTTGG 83315


RESULT 65
AC098121/c
LOCUS       AC098121               231321 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-133G18, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC098121
VERSION     AC098121.7  GI:30522725
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 231321)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
```

```
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 231321)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 231321)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24957732.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GGLN
                Center clone name: CH230-133G18
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 224443 bases at least Q40
                Consensus quality: 225706 bases at least Q30
                Consensus quality: 226616 bases at least Q20
                Estimated insert size: 232595; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   229838: contig of 229838 bp in length
            *  229839   229938: gap of unknown length
            *  229939   231321: contig of 1383 bp in length.
FEATURES             Location/Qualifiers
     source          1. .231321
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-133G18"
     misc_feature    1. .3236
                     /note="wgs_contig"
     misc_feature    227078. .228053
                     /note="wgs_contig"
     misc_feature    228104. .229838
                     /note="wgs_contig"
BASE COUNT    58426 a  49296 c  52671 g  67149 t   3779 others
ORIGIN

  Query Match            15.7%;  Score 157.6;  DB 2;  Length 231321;
  Best Local Similarity  53.3%;  Pred. No. 2.2e-16;
  Matches  358;  Conservative    0;  Mismatches  309;  Indels    5;  Gaps    1;

Qy        194 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 253
              ||  || |||   | |     |   ||||| ||| |||   |   | ||| | | ||
Db     182854 TGGTTGGTTGGCTGGCTGGTTGGCTGGCTGACTGGCTGGCTGGCTGGCTGGTGGTTGGTT 182795

Qy        254 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 313
              ||   || |||   ||  || | ||||   ||  |||   ||  || || |  ||  ||
Db     182794 GGTTGGTTGGCTGTCTGCCTGGTTGGCTGTCTGTCTGGTTGGTTTGTTGGTCGACTGCCT 182735

Qy        314 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 373
              |   ||   || ||  || | | | ||||||  ||| | ||   ||       ||   | |
Db     182734 GGCTGGCTGGTTGGTTTGTTGGTTGGCTGGCTGGCTGGTTGGCTGGCTGACTGGTTGGTT 182675

Qy        374 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 433
              || |  ||  | |    |||  |||| | || ||| ||| |||||   | |   ||| |
Db     182674 GGTTGACTGATTGTCTAGCTGACTGGTTTTCTGGCTGGCTGGCTGGTTGGTTGATTTGTT 182615

Qy        434 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 493
              | |||   ||| ||| || | ||||  ||| ||| | | | | |   | | ||| |||
Db     182614 GGTTGGTTGGCTGGCTGCCTGGTTGGCTGGCTGGCTGGTTGGTTTGTTGGTTGGCTGGCT 182555

Qy        494 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGG 553
              |||||   | ||| | | | ||||||  | | | | | |   ||||  ||| ||||
Db     182554 GGCTGGTTGGTTTGTTAGTTGGCTGGCTGGTTGGTTGGTTGATTGGCTGGCTGGCTGGCT 182495

Qy        554 GGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTA 613
              ||   || ||    || | |  |||| ||  |||   |  ||  | |||     | ||
Db     182494 GGTTGGTTGGTTGACTGGTTGTCTAACTGGCTGGTTTTCTGGCTAGCTGG-----TTGTC 182440

Qy        614 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 673
              |||||   || || | |   |  ||   ||||| | |   || || | |||||   |
Db     182439 TGGCTGGCTGGTTGATTGGTTGATTGGCTGGCTGGTTGATTGGTTGATTGGCTGGCTGGC 182380

Qy        674 GGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 733
              |||||   |||||   |  || |  | || || | |  || || | || || | |
Db     182379 TGGCTGGCTGGCTGGCTGATTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGT 182320

Qy        734 GGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 793
              || |  | || ||   |  |||||  |||||  | |  |||||  ||||  | |
Db     182319 TGGTTGGTTGGTTGGCTGGCTGGCTGACTGGCTGGTGGGCTGGCTGACTGGCTCTTTGGT 182260

Qy        794 GGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGC 853
              || || |   |  | |  | ||||||       ||||||  |||||  |   |
Db     182259 TGGTTGGTTTTGTTGTCTTTGCTATTCCTGCTATTGCTGTTGCTTATTTGCTTGCTTGCT 182200

Qy        854 CGCTTGCCTGCT 865
              |||  | | ||
Db     182199 TGCTCACTTTCT 182188


RESULT 66
AC124492
LOCUS       AC124492              176782 bp    DNA     linear   ROD 27-OCT-2002
DEFINITION  Mus musculus chromosome 12 clone RP23-454I23, complete sequence.
ACCESSION   AC124492
VERSION     AC124492.4  GI:24415450
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 176782)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176782)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 176782)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 176782)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Oct 27, 2002 this sequence version replaced gi:22758518.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ----------------- Project Information -----------------
            Center project name: M_BA0454I23
            ------------------------------------------------------.
FEATURES             Location/Qualifiers
     source          1. .176782
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="12"
                     /clone="RP23-454I23"
BASE COUNT    55394 a  35351 c  33704 g  52333 t
ORIGIN
```

Query Match 15.7%; Score 156.8; DB 10; Length 176782;
Best Local Similarity 53.2%; Pred. No. 3.2e-16;
Matches 332; Conservative 0; Mismatches 292; Indels 0; Gaps 0;

```
Qy        261 AGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGA 320
              | |||||| ||   |||  |||||  ||      | | |   |||||| |||      |
Db      40582 ATGCCTGCCTGTCTGCTTCCCTGCCCGCCTGCCTGCCTGCCTGCCCTGCCCGCCTGC 40641

Qy        321 CAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGC 380
              | |    ||||| ||   |||   ||||  ||      | | |   | |||   |||| |
Db      40642 CCGCTTCCCTGCCTGTCTGCTTCTCTGCCCGCCTGCCCGCCTGCCTGCCTGTCTGCTTCC 40701

Qy        381 TAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 440
                |   | |||| |||  ||| || |||  |||  |  |||||  ||||| |  ||  ||
Db      40702 CTGCCCGCCTGCCCGCCTGCCCGCCTGCCCGCTTCCCTGCCTGTCTGCTTCCCTGCCCGC 40761

Qy        441 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 500
              |||| ||| |||   ||  | ||||||  | ||| |||  ||| || |||  ||| ||  ||
Db      40762 CTGCCTGCCTGCCTGCCTGTCTGCTTCCCTGCCCGCCTGCCCGCCTGCCCGCTTGCCTGC 40821

Qy        501 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 560
              | | ||| |||  ||  ||||||  ||||||  |||| ||||||||  |      ||| |
Db      40822 TTCCCTGCCTGCCTGCCCGCCTGCCCGCTTGCCCGCTTGCCTGCTTCCCTGCCCGACTGC 40881

Qy        561 AGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGC 620
                |  |||  |||| |      | | |  ||||||||  |   |       |   | ||||
Db      40882 CCGCTTGCCCGCTTTCCTGCCCGCCTGCCCGCCTGCCCGCCTGACCACCTGCCCGCCTGC 40941

Qy        621 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 680
              | ||||  |   | | ||  | | ||  |     ||||  |  |||  |       ||||
Db      40942 CCGCCTGCCCGCCTGCCCGCCTGCCCGCCTGCCCACCTGCCCGCTTGCCTGCTTCCCTGC 41001

Qy        681 TGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCC 740
```

```
                |||| |   | |  || | |||||     | | ||  | | ||| |   |  |||
Db      41002 CCACCTGCCCGCCTGCCCGCTTGCCTGCTTCCCTGCCCGCCTGCCCGCTTGCCTGCTTCC 41061

Qy        741 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 800
                | ||  |     ||||    ||||       | |  | | | |  |     ||||
Db      41062 CTACCCGCCTGCTCACCTGCCCACCTGCCCACCTGCCCACCTGCCCACCTGCCCACCTGC 41121

Qy        801 TGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGC 860
                |||| |   | | ||||||  |||| | ||||| | |||      | | || |||
Db      41122 CCACCTGCCCGCCTGTCCGCCTGCCCGCTTCCCTGCTTCCCTGCCCACCTGCCTGCCTGC 41181

Qy        861 CTGCTACTAGGGCTGCTGTGCAGC 884
                | |      | |||||  | | ||
Db      41182 TTCCCTGCCTGCCTGCTTTCCTGC 41205
```

```
RESULT 67
AC142424/c
LOCUS       AC142424              140898 bp    DNA     linear   HTG 01-APR-2003
DEFINITION  Rattus norvegicus clone CH230-408K24, WORKING DRAFT SEQUENCE, 39
            unordered pieces.
ACCESSION   AC142424
VERSION     AC142424.2  GI:29423803
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 140898)
  AUTHORS   Muzny,D.Marie.,,Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A.,
            Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
            Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
            Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G.,
            Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
            Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
            Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
            Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
            Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
            Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
            Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
            Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
            Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
            Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
            Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
            Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
```

```
            Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
            Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
            Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
            Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
            Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 140898)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 140898)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Apr 1, 2003 this sequence version replaced gi:29366848.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KEPT
                Center clone name: CH230-408K24
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 126303 bases at least Q40
                Consensus quality: 130569 bases at least Q30
                Consensus quality: 134257 bases at least Q20
                Estimated insert size: 128743; sum-of-contigs estimation
                Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 39 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1198: contig of 1198 bp in length
            *     1199     1298: gap of unknown length
            *     1299     2537: contig of 1239 bp in length
            *     2538     2637: gap of unknown length
            *     2638     3734: contig of 1097 bp in length
            *     3735     3834: gap of unknown length
            *     3835     5472: contig of 1638 bp in length
            *     5473     5572: gap of unknown length
            *     5573     7555: contig of 1983 bp in length
            *     7556     7655: gap of unknown length
            *     7656     8798: contig of 1143 bp in length
            *     8799     8898: gap of unknown length
            *     8899    10417: contig of 1519 bp in length
            *    10418    10517: gap of unknown length
            *    10518    11625: contig of 1108 bp in length
            *    11626    11725: gap of unknown length
            *    11726    12855: contig of 1130 bp in length
            *    12856    12955: gap of unknown length
            *    12956    14751: contig of 1796 bp in length
            *    14752    14851: gap of unknown length
            *    14852    16352: contig of 1501 bp in length
            *    16353    16452: gap of unknown length
            *    16453    19169: contig of 2717 bp in length
            *    19170    19269: gap of unknown length
            *    19270    21464: contig of 2195 bp in length
            *    21465    21564: gap of unknown length
            *    21565    22701: contig of 1137 bp in length
            *    22702    22801: gap of unknown length
            *    22802    24203: contig of 1402 bp in length
            *    24204    24303: gap of unknown length
            *    24304    26258: contig of 1955 bp in length
            *    26259    26358: gap of unknown length
            *    26359    28665: contig of 2307 bp in length
            *    28666    28765: gap of unknown length
            *    28766    30207: contig of 1442 bp in length
            *    30208    30307: gap of unknown length
            *    30308    32693: contig of 2386 bp in length
            *    32694    32793: gap of unknown length
            *    32794    34145: contig of 1352 bp in length
            *    34146    34245: gap of unknown length
            *    34246    35342: contig of 1097 bp in length
            *    35343    35442: gap of unknown length
            *    35443    37116: contig of 1674 bp in length
            *    37117    37216: gap of unknown length
            *    37217    38848: contig of 1632 bp in length
            *    38849    38948: gap of unknown length
            *    38949    42968: contig of 4020 bp in length
            *    42969    43068: gap of unknown length
            *    43069    46012: contig of 2944 bp in length
            *    46013    46112: gap of unknown length
            *    46113    51041: contig of 4929 bp in length
            *    51042    51141: gap of unknown length
            *    51142    55565: contig of 4424 bp in length
            *    55566    55665: gap of unknown length
            *    55666    59317: contig of 3652 bp in length
            *    59318    59417: gap of unknown length
            *    59418    63593: contig of 4176 bp in length
            *    63594    63693: gap of unknown length
            *    63694    67952: contig of 4259 bp in length
            *    67953    68052: gap of unknown length
            *    68053    73303: contig of 5251 bp in length
            *    73304    73403: gap of unknown length
            *    73404    78136: contig of 4733 bp in length
            *    78137    78236: gap of unknown length
            *    78237    82228: contig of 3992 bp in length
            *    82229    82328: gap of unknown length
            *    82329    88954: contig of 6626 bp in length
            *    88955    89054: gap of unknown length
            *    89055    96366: contig of 7312 bp in length
            *    96367    96466: gap of unknown length
            *    96467   103174: contig of 6708 bp in length
            *   103175   103274: gap of unknown length
            *   103275   113341: contig of 10067 bp in length
            *   113342   113441: gap of unknown length
            *   113442   126269: contig of 12828 bp in length
            *   126270   126369: gap of unknown length
            *   126370   140898: contig of 14529 bp in length.
FEATURES             Location/Qualifiers
     source          1. .140898
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-408K24"
BASE COUNT    37480 a  28929 c  29223 g  41432 t   3834 others
ORIGIN

  Query Match           15.6%;  Score 156.6;  DB 2;  Length 140898;
  Best Local Similarity   53.8%;  Pred. No. 3.6e-16;
  Matches  348;  Conservative    0;  Mismatches  294;  Indels    5;  Gaps    1;

Qy        219 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 278
              ||||| ||| |||   |   | ||| | | ||   ||   || |||   ||  || | ||
Db      45993 GGCTGACTGGCTGGCTGGCTGGTGGTTGGTTGGTTGGTTGGCTGTCTGCCTGGTTG 45934

Qy        279 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 338
              ||   ||  |||   ||   || || |  ||  |||   ||   || ||  || | | |
Db      45933 GCTGTCTGTCTGGTTGGTTTGTTGGTCGACTGCCTGGCTGGCTGGTTGGTTTGTTGGTTG 45874
```

```
Qy        339 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 398
              ||||||  ||| | ||   ||       ||   | ||| |  ||  | |    |||  |||
Db      45873 GCTGGCTGGCTGGTTGGCTGGCTGACTGGTTGGTTGGTTGACTGATTGTCTAGCTGACTG 45814

Qy        399 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 458
              |  | || ||| ||| |||||   | |   ||| | | |||   ||| |||  || | ||
Db      45813 GTTTTCTGGCTGGCTGGCTGGTTGGTTGATTTGTTGGTTGGTTGGCTGGCTGCCTGGTTG 45754

Qy        459 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 518
              ||  ||| ||| | | | | |   | | ||| ||| |||||   | ||| | | | ||||
Db      45753 GCTGGCTGGCTGGTTGGTTTGTTGGTTGGCTGGCTGGCTGGTTGGTTTGTTAGTTGGCTG 45694

Qy        519 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTA 578
              ||  | | | | | |   ||||  ||| ||||   ||   || ||    || | |  |||
Db      45693 GCTGGTTGGTTGGTTGATTGGCTGGCTGGCTGGCTGGTTGGTTGGTTGACTGGTTGTCTA 45634

Qy        579 AGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 638
              |  ||  |||   |  ||  | ||| |     ||  |||||   || || | |   |  |
Db      45633 ACTGGCTGGTTTTCTGGCTAGCTGGTT-----GTCTGGCTGGCTGGTTGATTGGTTGATT 45579

Qy        639 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 698
              |   ||||| | |   || || | |||||   |   |||||  |||||   |   || |
Db      45578 GGCTGGCTGGTTGATTGGTTGATTGGCTGGCTGGCTGGCTGGCTGGCTGATTGGTT 45519

Qy        699 CCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCT 758
              | || || | |   || || | || || | |   || |  | || ||   |  ||||
Db      45518 GGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGGCT 45459

Qy        759 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 818
              |  |||||  | |  |||||  ||||  | |  || || |    |  |  | |  |
Db      45458 GACTGGCTGGTGGGCTGGCTGACTGGCTCTTTGGTTGGTTGGTTTTGTTGTCTTTGCTAT 45399

Qy        819 GCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCT 865
              ||||||       ||||||  |||||  |   |   |||  | | ||
Db      45398 TCCTGCTATTGCTGTTGCTTATTTGCTTGCTTGCTTGCTCACTTTCT 45352
```

```
RESULT 68
AC044848
LOCUS       AC044848              179192 bp    DNA     linear   HTG 09-MAY-2001
DEFINITION  Mus musculus chromosome 11 clone RP23-129L22 map 11, WORKING DRAFT
            SEQUENCE, 70 unordered pieces.
ACCESSION   AC044848
VERSION     AC044848.3  GI:12658164
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 179192)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 11, clone RP23-129L22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179192)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 3, 2001 this sequence version replaced gi:10717286.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9178
                Center clone name: 129_L_22
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 7% of reads
                Sequencing vector: Plasmid; n/a; 93% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 151306 bases at least Q40
                Consensus quality: 164443 bases at least Q30
                Consensus quality: 169708 bases at least Q20
                Insert size: 230000; agarose-fp
                Insert size: 172292; sum-of-contigs
                Quality coverage: 2.4 in Q20 bases; agarose-fp
                Quality coverage: 3.2 in Q20 ba.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 70 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      399: contig of 399 bp in length
            *      400      499: gap of 100 bp
            *      500     1142: contig of 643 bp in length
            *     1143     1242: gap of 100 bp
            *     1243     1653: contig of 411 bp in length
            *     1654     1753: gap of 100 bp
            *     1754     2376: contig of 623 bp in length
            *     2377     2476: gap of 100 bp
            *     2477     3095: contig of 619 bp in length
            *     3096     3195: gap of 100 bp
            *     3196     4176: contig of 981 bp in length
            *     4177     4276: gap of 100 bp
            *     4277     5264: contig of 988 bp in length
            *     5265     5364: gap of 100 bp
            *     5365     6039: contig of 675 bp in length
            *     6040     6139: gap of 100 bp
            *     6140     6817: contig of 678 bp in length
            *     6818     6917: gap of 100 bp
            *     6918     7954: contig of 1037 bp in length
            *     7955     8054: gap of 100 bp
            *     8055     8811: contig of 757 bp in length
            *     8812     8911: gap of 100 bp
            *     8912     9775: contig of 864 bp in length
            *     9776     9875: gap of 100 bp
            *     9876    10921: contig of 1046 bp in length
            *    10922    11021: gap of 100 bp
            *    11022    12232: contig of 1211 bp in length
            *    12233    12332: gap of 100 bp
            *    12333    13351: contig of 1019 bp in length
            *    13352    13451: gap of 100 bp
            *    13452    14326: contig of 875 bp in length
            *    14327    14426: gap of 100 bp
            *    14427    15792: contig of 1366 bp in length
            *    15793    15892: gap of 100 bp
            *    15893    16562: contig of 670 bp in length
            *    16563    16662: gap of 100 bp
```

```
*    16663    17540: contig of 878 bp in length
*    17541    17640: gap of 100 bp
*    17641    18844: contig of 1204 bp in length
*    18845    18944: gap of 100 bp
*    18945    20507: contig of 1563 bp in length
*    20508    20607: gap of 100 bp
*    20608    21903: contig of 1296 bp in length
*    21904    22003: gap of 100 bp
*    22004    23064: contig of 1061 bp in length
*    23065    23164: gap of 100 bp
*    23165    24725: contig of 1561 bp in length
*    24726    24825: gap of 100 bp
*    24826    26288: contig of 1463 bp in length
*    26289    26388: gap of 100 bp
*    26389    28040: contig of 1652 bp in length
*    28041    28140: gap of 100 bp
*    28141    28975: contig of 835 bp in length
*    28976    29075: gap of 100 bp
*    29076    29960: contig of 885 bp in length
*    29961    30060: gap of 100 bp
*    30061    31316: contig of 1256 bp in length
*    31317    31416: gap of 100 bp
*    31417    32758: contig of 1342 bp in length
*    32759    32858: gap of 100 bp
*    32859    34220: contig of 1362 bp in length
*    34221    34320: gap of 100 bp
*    34321    35754: contig of 1434 bp in length
*    35755    35854: gap of 100 bp
*    35855    37185: contig of 1331 bp in length
*    37186    37285: gap of 100 bp
*    37286    39443: contig of 2158 bp in length
*    39444    39543: gap of 100 bp
*    39544    41045: contig of 1502 bp in length
*    41046    41145: gap of 100 bp
*    41146    42858: contig of 1713 bp in length
*    42859    42958: gap of 100 bp
*    42959    44585: contig of 1627 bp in length
*    44586    44685: gap of 100 bp
*    44686    46869: contig of 2184 bp in length
*    46870    46969: gap of 100 bp
*    46970    48860: contig of 1891 bp in length
*    48861    48960: gap of 100 bp
*    48961    51281: contig of 2321 bp in length
*    51282    51381: gap of 100 bp
*    51382    53943: contig of 2562 bp in length
*    53944    54043: gap of 100 bp
*    54044    56936: contig of 2893 bp in length
*    56937    57036: gap of 100 bp
*    57037    59918: contig of 2882 bp in length
*    59919    60018: gap of 100 bp
*    60019    74665: contig of 14647 bp in length
*    74666    74765: gap of 100 bp
*    74766    76556: contig of 1791 bp in length
*    76557    76656: gap of 100 bp
*    76657    78619: contig of 1963 bp in length
*    78620    78719: gap of 100 bp
*    78720    80541: contig of 1822 bp in length
*    80542    80641: gap of 100 bp
*    80642    83536: contig of 2895 bp in length
*    83537    83636: gap of 100 bp
*    83637    87151: contig of 3515 bp in length
*    87152    87251: gap of 100 bp
*    87252    89340: contig of 2089 bp in length
*    89341    89440: gap of 100 bp
*    89441    91459: contig of 2019 bp in length
*    91460    91559: gap of 100 bp
*    91560    93525: contig of 1966 bp in length
*    93526    93625: gap of 100 bp
*    93626    95801: contig of 2176 bp in length
*    95802    95901: gap of 100 bp
*    95902    98253: contig of 2352 bp in length
*    98254    98353: gap of 100 bp
*    98354   100651: contig of 2298 bp in length
*   100652   100751: gap of 100 bp
*   100752   103409: contig of 2658 bp in length
*   103410   103509: gap of 100 bp
*   103510   107033: contig of 3524 bp in length
*   107034   107133: gap of 100 bp
*   107134   110878: contig of 3745 bp in length
*   110879   110978: gap of 100 bp
*   110979   113971: contig of 2993 bp in length
*   113972   114071: gap of 100 bp
*   114072   117157: contig of 3086 bp in length
*   117158   117257: gap of 100 bp
*   117258   120573: contig of 3316 bp in length
*   120574   120673: gap of 100 bp
*   120674   123933: contig of 3260 bp in length
*   123934   124033: gap of 100 bp
*   124034   129503: contig of 5470 bp in length
*   129504   129603: gap of 100 bp
*   129604   134903: contig of 5300 bp in length
*   134904   135003: gap of 100 bp
*   135004   141319: contig of 6316 bp in length

  Query Match             15.6%;  Score 156.6;  DB 2;  Length 179192;
  Best Local Similarity   59.2%;  Pred. No. 3.4e-16;
  Matches  267;  Conservative    0;  Mismatches  184;  Indels    0;  Gaps    0;

Qy        219 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 278
              | ||||||||||||  ||  |||||| |||         ||| | |||||| |||  ||
Db      88631 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCT 88690

Qy        279 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 338
              ||||||  | |       | | |   |||||| |||       | | |   |||||| |||
Db      88691 GCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 88750

Qy        339 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 398
              ||  ||||  |  ||       | | |   | ||||   |  ||  |   | ||||  ||
Db      88751 GCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCT 88810

Qy        399 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 458
              ||||||  |||  |||  ||  |||||| |||  | |  ||  |||||| |||  |||  ||
Db      88811 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCT 88870

Qy        459 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 518
              ||||  |  ||  |||  ||  |||||| |||  |||  ||  |||||| |||  |  ||
Db      88871 GCCTCCCAGCATGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCAT 88930

Qy        519 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTA 578
              ||||||  |||  |||  ||  ||||||  |||       | | |   | ||||  |||  |  |
Db      88931 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCA 88990

Qy        579 AGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 638
                | | |   |||||||||   || |   || | | ||||  | ||||  |   | ||
Db      88991 GCATGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTGCCT 89050

Qy        639 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 669
              ||  | ||||  |   | ||||  |  |||||
Db      89051 GCCTGCCTGCCTGCCTGCCTGCAGCACTGCT 89081


RESULT 69
AL583887/c
LOCUS       AL583887               220050 bp    DNA     linear   ROD 09-OCT-2001
DEFINITION  Mouse DNA sequence from clone RP23-121M7 on chromosome 15, complete
            sequence.
ACCESSION   AL583887
VERSION     AL583887.9  GI:16030170
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 220050)
  AUTHORS   Phillimore,B.
  TITLE     Direct Submission
```

```
JOURNAL     Submitted (09-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 10, 2001 this sequence version replaced gi:14269983.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-121M7 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP23-121M7 The true
            left end of clone RP23-85M21 is at 142216 in this sequence. The
            true right end of clone RP23-10K12 is at 78004 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .220050
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="15"
                     /clone="RP23-121M7"
                     /clone_lib="RPCI-23"
     repeat_region   28. .453
                     /note="RMER10A repeat: matches 12. .426 of consensus"
     misc_feature    304. .316
                     /note="Sequence from overlapping clone RP23-10K12
                     (AL583889).Assembly confirmed by restriction digest."
     repeat_region   864. .893
                     /note="5 copies 6 mer gctggg 93% conserved"
     repeat_region   873. .968
                     /note="2 copies 48 mer 94% conserved"
     repeat_region   896. .979
                     /note="14 copies 6 mer ggctgg 67% conserved"
     repeat_region   1006. .1123
                     /note="B1_MM repeat: matches 71. .146 of consensus"
     repeat_region   1534. .1680
                     /note="B1_MM repeat: matches 1. .146 of consensus"
     repeat_region   1681. .1788
                     /note="9 copies 12 mer 80% conserved"
     repeat_region   1682. .1801
                     /note="3 copies 40 mer 80% conserved"
     repeat_region   1690. .1793
                     /note="4 copies 26 mer 79% conserved"
     repeat_region   1789. .1933
                     /note="29 copies 5 mer agggg 62% conserved"
     repeat_region   1795. .1938
                     /note="3 copies 48 mer 80% conserved"
     repeat_region   1796. .1933
                     /note="69 copies 2 mer gg 60% conserved"
     repeat_region   1802. .1927
                     /note="21 copies 6 mer gagggg 63% conserved"
     repeat_region   1806. .1935
                     /note="5 copies 26 mer 69% conserved"
     repeat_region   1814. .1930
                     /note="3 copies 39 mer 76% conserved"
     repeat_region   1840. .1923
                     /note="21 copies 4 mer gagg 65% conserved"
     repeat_region   1962. .2129
                     /note="B3A repeat: matches 1. .192 of consensus"
     repeat_region   2130. .2266
                     /note="B1_MM repeat: matches 1. .146 of consensus"
     gene            complement(2267. .29223)
                     /gene="bM121M7.1"
     mRNA            complement(join(2267. .2680,7836. .7999,10910. .10996,
                     14639. .14782,15687. .15795,16519. .16653,18289. .18469,
                     20448. .20656,23617. .23733,23850. .24010,24647. .24694,
                     29101. .29223))
                     /gene="bM121M7.1"
                     /product="bM121M7.1 (novel protein (ortholog of human
                     tubulin tyrosine ligase-like 1 (TTLL1)), isoform 1)"
                     /note="match: cDNAs: Em:BC009058 Em:AF104927 Em:AL096883
                     Em:BC010510 Em:AL136687 Em:AF173935 Em:AL096886
                     match: ESTs: Em:BI359712 Em:BG895852 Em:BF567059
                     Em:AW657089 Em:AW901129 Em:BE651256 Em:BI181718
                     Em:BI084551"
                     /evidence=not_experimental
     polyA_site      complement(2267)
                     /gene="bM121M7.1"
     misc_feature    complement(2283. .2632)
                     /gene="bM121M7.1"
                     /note="match: GSS: Em:AZ777040"
     repeat_region   2304. .2371
                     /note="17 copies 4 mer gtgt 80% conserved"
     repeat_region   2305. .2382
                     /note="3 copies 26 mer 79% conserved"
     repeat_region   2311. .2370
                     /note="5 copies 12 mer 88% conserved"
     polyA_signal    complement(2426. .2431)
                     /gene="bM121M7.1"
     CDS             complement(join(2551. .2680,7836. .7999,14639. .14782,
                     15687. .15795,16519. .16653,18289. .18469,20448. .20656,
                     23617. .23729))
                     /gene="bM121M7.1"
                     /note="match: proteins: Tr:Q9BR27"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="bM121M7.2 (novel protein (ortholog of human
                     tubulin tyrosine ligase-like 1 (TTLL1)), isoform 2)"
                     /protein_id="CAC93920.1"
                     /db_xref="GI:16030171"
                     /translation="MAGRVKWVTDIEKSVLINNFEKRGWIQVTENEDWNFYWMSVQTI
                     RNVFSVETGYRLSDDQIVNHFPNHYELTRKDLMVKNIKRYRKELEKEGSPLAEKDENG
                     KYLYLDFVPVTYMLPADYNLFVEEFRKSPSSTWIMKPCGKAQGKGIFLINKLSQIKKW
                     SRDSKTSSFVSQSTKEAYVISVYINNPLLIGGRKFDLRLYVLVSTYRPLRCYMYKLGF
                     CRFCTVKYTPSTSELDNMFVHLTNVAIQKHGEDYNHIHGGKWTVNNLRLYLESTRGRE
                     VTSKLFDEIHWIIVQSLKAVAVNASPSLTSSTANDRILKYNLINDTLNIAVPNGEIPD
                     CKWNKSPPKEVLGNYEILYDEELAQGDGAERELRNRPGQPVGPRAGRSRDSGRSVLTT
                     WK"
     CDS             complement(join(2551. .2680,7836. .7999,10910. .10996,
                     14639. .14782,15687. .15795,16519. .16653,18289. .18469,
                     20448. .20656,23617. .23729))
                     /gene="bM121M7.1"
                     /note="match: proteins: Tr:O95922 Tr:Q9UMU0 Tr:AAH10510
                     Tr:AAK92989 Tr:Q9VZ97 Tr:Q9VX74"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="bM121M7.1 (novel protein (ortholog of human
                     tubulin tyrosine ligase-like 1 (TTLL1)), isoform 1)"
                     /protein_id="CAC93921.1"
                     /db_xref="GI:16030172"
                     /translation="MAGRVKWVTDIEKSVLINNFEKRGWIQVTENEDWNFYWMSVQTI
                     RNVFSVETGYRLSDDQIVNHFPNHYELTRKDLMVKNIKRYRKELEKEGSPLAEKDENG
                     KYLYLDFVPVTYMLPADYNLFVEEFRKSPSSTWIMKPCGKAQGKGIFLINKLSQIKKW
                     SRDSKTSSFVSQSTKEAYVISVYINNPLLIGGRKFDLRLYVLVSTYRPLRCYMYKLGF
                     CRFCTVKYTPSTSELDNMFVHLTNVAIQKHGEDYNHIHGGKWTVNNLRLYLESTRGRE
                     VTSKLFDEIHWIIVQSLKAVAPVMNNDKHCFECYGYDIIIDDKLKPWLIEVNASPSLT
                     SSTANDRILKYNLINDTLNIAVPNGEIPDCKWNKSPPKEVLGNYEILYDEELAQGDGA
                     ERELRNRPGQPVGPRAGRSRDSGRSVLTTWK"
     repeat_region   2983. .3138
                     /note="B2_Mm2 repeat: matches 1. .157 of consensus"
     repeat_region   3147. .3328
```

```
                     /note="7 copies 26 mer 63% conserved"
     repeat_region   3164. .3358
                     /note="5 copies 39 mer 69% conserved"
     repeat_region   3219. .3302
                     /note="3 copies 28 mer 79% conserved"
     repeat_region   3256. .3353
                     /note="49 copies 2 mer aa 67% conserved"
     repeat_region   3269. .3352
                     /note="7 copies 12 mer 71% conserved"
     misc_feature    complement(3536. .4243)
                     /gene="bM121M7.1"
                     /note="match: GSS: Em:AQ975102"
     repeat_region   3792. .3919
                     /note="RSINE1 repeat: matches 4. .131 of consensus"
     repeat_region   3920. .4092
                     /note="RMER12 repeat: matches 1. .170 of consensus"
     repeat_region   4161. .4262
                     /note="RMER12 repeat: matches 177. .297 of consensus"
     repeat_region   4287. .4575
                     /note="RMER17C repeat: matches 114. .414 of consensus"
     repeat_region   4576. .5014
                     /note="RMER12 repeat: matches 289. .743 of consensus"
     repeat_region   5031. .5080
                     /note="L1MC4 repeat: matches 6656. .6703 of consensus"
     repeat_region   5154. .5183
                     /note="5 copies 6 mer gtgtgt 96% conserved"
     repeat_region   5239. .5369
                     /note="RMER12 repeat: matches 1185. .1315 of consensus"
     repeat_region   5422. .5555
                     /note="RSINE1 repeat: matches 3. .149 of consensus"
     repeat_region   5556. .5682
                     /note="B1_MM repeat: matches 1. .125 of consensus"
     repeat_region   5713. .5824
                     /note="B1_MM repeat: matches 1. .112 of consensus"
     misc_feature    5869. .6327
                     /note="match: GSS: Em:AZ713136"
     repeat_region   6388. .6421
                     /note="17 copies 2 mer ac 100% conserved"
     repeat_region   6812. .6824
                     /note="B4 repeat: matches 74. .86 of consensus"
     repeat_region   6825. .6931

  Query Match             15.6%;  Score 156.6;  DB 10;  Length 220050;
  Best Local Similarity   59.2%;  Pred. No. 3.2e-16;
  Matches  267;  Conservative    0;  Mismatches  184;  Indels    0;  Gaps    0;

Qy        219 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 278
              | ||||||||||||  ||  ||||||  |||        |||  | ||||||  |||  ||
Db      22928 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCT 22869

Qy        279 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 338
              ||||||  | |       | | |   ||||||  |||       | | |   ||||||  |||
Db      22868 GCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 22809

Qy        339 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 398
              ||  ||||  |  ||       | | |   | ||||   |  ||  |   | ||||  ||
Db      22808 GCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCT 22749

Qy        399 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 458
              ||||||  |||  |||  ||  ||||||  |||  |  |  ||  ||||||  |||  |||  ||
Db      22748 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTGCCT 22689

Qy        459 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 518
              ||||  |  ||  |||  ||  ||||||  |||  |||  ||  ||||||  |||  |  |  ||
Db      22688 GCCTCCCAGCATGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCAT 22629

Qy        519 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTA 578
              ||||||  |||  |||  ||  ||||||  |||       | | |   | ||||  |||  | | |
Db      22628 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCA 22569

Qy        579 AGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 638
                | | |   |||||||||   ||  |   ||  | |  ||||   | ||||   |   | ||
Db      22568 GCATGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTGCCTGCCT 22509

Qy        639 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 669
              ||  | ||||  |   | ||||  |  |||||
Db      22508 GCCTGCCTGCCTGCCTGCCTGCAGCACTGCT 22478


RESULT 70
AL583889
LOCUS       AL583889              236344 bp    DNA     linear   ROD 05-APR-2002
DEFINITION  Mouse DNA sequence from clone RP23-10K12 on chromosome 15, complete
            sequence.
ACCESSION   AL583889
VERSION     AL583889.8  GI:16944094
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Tromans,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 15, 2001 this sequence version replaced gi:16596742.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-10K12 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .236344
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="15"
                     /clone="RP23-10K12"
                     /clone_lib="RPCI-23"
BASE COUNT    57826 a  55744 c  58916 g  63858 t
ORIGIN

  Query Match             15.6%;  Score 156.6;  DB 10;  Length 236344;
  Best Local Similarity   59.2%;  Pred. No. 3.2e-16;
  Matches  267;  Conservative    0;  Mismatches  184;  Indels    0;  Gaps    0;

Qy        219 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 278
              | ||||||||||||  ||  ||||||  |||        |||  | ||||||  |||  ||
Db      55077 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCT 55136

Qy        279 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 338
              ||||||  | |       | | |   ||||||  |||       | | |   ||||||  |||
Db      55137 GCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 55196

Qy        339 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 398
              ||  ||||  |  ||       | | |   | ||||   |  ||  |   | ||||  ||
Db      55197 GCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCT 55256
```

```
Qy        399 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 458
              |||||| ||| |||  ||  |||||| ||| | |  ||  |||||| ||| |||  ||
Db      55257 GCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTGCCT 55316

Qy        459 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 518
              |||| |  || |||  ||  |||||| ||| ||| ||  |||||| ||| | |  ||
Db      55317 GCCTCCCAGCATGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCAGCAT 55376

Qy        519 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTA 578
              |||||| ||| |||  ||  |||||| |||       | | |   | |||| ||| | | |
Db      55377 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCA 55436

Qy        579 AGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 638
                  | | |   |||||||||   || |   || | | ||||  | ||||  |   | ||
Db      55437 GCATGCCTGCCTGCCTGCCTGCCTCCCAGCATGCCTGCCTGCCTGCCTGCCTGCCT 55496

Qy        639 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 669
              ||  | ||||  |   | |||| |  |||||
Db      55497 GCCTGCCTGCCTGCCTGCCTGCAGCACTGCT 55527
```

RESULT 71
AC106240/c
LOCUS AC106240 251656 bp DNA linear HTG 22-SEP-2002
DEFINITION Rattus norvegicus clone CH230-129E1, *** SEQUENCING IN PROGRESS ***, 6 unordered pieces.
ACCESSION AC106240
VERSION AC106240.3 GI:23268191
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 251656)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 251656)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 251656)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (22-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Sep 22, 2002 this sequence version replaced gi:21731477.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequening reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
-------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
Center-project name: GKEV
Center clone name: CH230-129E1
-------------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 226971 bases at least Q40
Consensus quality: 228331 bases at least Q30
Consensus quality: 229024 bases at least Q20
Estimated insert size: 260341; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 71565: contig of 71565 bp in length
* 71566 71665: gap of unknown length
* 71666 165635: contig of 93970 bp in length
* 165636 165735: gap of unknown length
* 165736 246680: contig of 80945 bp in length
* 246681 246780: gap of unknown length
* 246781 247856: contig of 1076 bp in length
* 247857 247956: gap of unknown length
* 247957 249454: contig of 1498 bp in length
* 249455 249554: gap of unknown length
* 249555 251656: contig of 2102 bp in length.
FEATURES Location/Qualifiers

```
     source          1. .251656
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-129E1"
     misc_feature    1. .1121
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    21157. .21283
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH286379"
     misc_feature    71666. .73022
                     /note="wgs_contig"
BASE COUNT    62675 a  52099 c  51098 g  63549 t  22235 others
ORIGIN

  Query Match             15.6%;  Score 156.4;  DB 2;  Length 251656;
  Best Local Similarity   55.0%;  Pred. No. 3.3e-16;
  Matches  349;  Conservative    0;  Mismatches  281;  Indels    4;  Gaps    2;

Qy        233 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAG 292
              | | |||   |   | ||   |    || ||  || ||| | | |||  || ||| | |
Db     223684 TGGTTGGTTGGTGGGTTGGTTGATTGGTGGGGTTGGTTGGTTGGTGGATTGGTTGGTTGG 223625

Qy        293 GGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGC 352
               |    |   |  || ||| |  | |   |   |  || | | | | ||| || |||
Db     223624 TGTGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTGGATTGGTGGGTTGGTTGG 223565

Qy        353 TGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGC 412
              || |  |   |  ||| || | | ||| | |   || || |   |||  || ||| |||
Db     223564 TGGGTTGGTTGGTGGGTTGGTTGGTTGGTGGGTTGGTTGGTTTGTGGGTTGGTTGGTTGG 223505

Qy        413 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 472
              | | |||  || ||| | | | | | |  || ||| | | | | |||  || |||  ||
Db     223504 TTGGTGGGTTGGTTGGTGGGTTGGTTGGTTGGTTGGTGGGTTGGTGGGTTGGTTGGATGG 223445

Qy        473 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 532
              | | |||  || ||| ||| | | | |  || ||| | | | | |||  || ||| | |
Db     223444 TTGGTGGGTTGGTTGGTTGGTGGGTTGGTTGGTTGGTGGGTTGGTGGGTTGGTTGGTGGG 223385

Qy        533 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAG-GGGACGGTACG 591
              | | |||  || ||| || | |   |   || || ||| ||| | |  || ||| ||
Db     223384 TTGGTGGGTTGGTTGGTGGGTTGGTTGGTTGGTTGGTTGGTTGGTGGGTTGGTTGGTTCG 223325

Qy        592 CCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAG 651
                |   | ||| ||| | |  ||| || |||| || | |  ||| || | || || | |
Db     223324 TTGGTGGGTTGGTTTGGTTGGTGGGTTGGTGGGTTGGTTGGTGGGTTGGTTGGTTGGTTG 223265

Qy        652 TAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAG 711
              || || |||| || | |   ||  | | || || | |  ||| |  | ||  |   |
Db     223264 GTTGGTTGGTGGGTTGGTGGGTTGGTGGGTTGGTTGGTTGGTGGGTTGGTTGGTGG---G 223208

Qy        712 TAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAG 771
              | ||  || | || || | |   || |  | || || | |  ||| || | || || | |
Db     223207 TTGGTTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTGGGTTGGTTGGTTGGTTG 223148

Qy        772 TAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTG 831
                || || | || || | |  ||| || | || || | |   || ||   | | | | |
Db     223147 GTTGGTTGGTTGGTTGATTGGTGGGTTGGTTGGTTGGTGGATTGGTTGGTGGGTTGGTGG 223088

Qy        832 CTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCT 865
               ||| ||| ||| | |  |||  | |||  || |
Db     223087 GTTGGTTGATTGGTTGGTGGGTTGGTTGGTTGAT 223054


RESULT 72
AC122561/c
LOCUS       AC122561               208994 bp    DNA     linear   HTG 17-MAR-2003
DEFINITION  Mus musculus clone RP24-97G2, WORKING DRAFT SEQUENCE, 10 unordered
            pieces.
ACCESSION   AC122561
VERSION     AC122561.3  GI:28975915
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 208994)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-97G2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 208994)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 208994)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 17, 2003 this sequence version replaced gi:28882445.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L26523
                Center clone name: 97_G_2
            -------------- Summary Statistics
```

```
            Sequencing vector: Plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 205816 bases at least Q40
            Consensus quality: 206724 bases at least Q30
            Consensus quality: 207402 bases at least Q20
            Insert size: 183000; agarose-fp
            Insert size: 208094; sum-of-contigs
            Quality coverage: 9.8 in Q20 bases; agarose-fp
            Quality coverage: 8.7 in Q20 bases; sum-of-contigs
            ----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2678: contig of 2678 bp in length
            *     2679     2778: gap of 100 bp
            *     2779     7560: contig of 4782 bp in length
            *     7561     7660: gap of 100 bp
            *     7661    17565: contig of 9905 bp in length
            *    17566    17665: gap of 100 bp
            *    17666    31160: contig of 13495 bp in length
            *    31161    31260: gap of 100 bp
            *    31261    53826: contig of 22566 bp in length
            *    53827    53926: gap of 100 bp
            *    53927    84289: contig of 30363 bp in length
            *    84290    84389: gap of 100 bp
            *    84390   109590: contig of 25201 bp in length
            *   109591   109690: gap of 100 bp
            *   109691   140213: contig of 30523 bp in length
            *   140214   140313: gap of 100 bp
            *   140314   172070: contig of 31757 bp in length
            *   172071   172170: gap of 100 bp
            *   172171   208994: contig of 36824 bp in length.
FEATURES             Location/Qualifiers
     source          1. .208994
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-97G2"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .2678
                     /note="assembly_fragment"
     misc_feature    2779. .7560
                     /note="assembly_fragment"
     misc_feature    7661. .17565
                     /note="assembly_fragment"
     misc_feature    17666. .31160
                     /note="assembly_fragment"
     misc_feature    31261. .53826
                     /note="assembly_fragment"
     misc_feature    53927. .84289
                     /note="assembly_fragment"
     misc_feature    84390. .109590
                     /note="assembly_fragment"
     misc_feature    109691. .140213
                     /note="assembly_fragment"
     misc_feature    140314. .172070
                     /note="assembly_fragment"
     misc_feature    172171. .208994
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    61737 a  42018 c  43038 g  61299 t    902 others
ORIGIN

  Query Match           15.5%;  Score 155.4;  DB 2;  Length 208994;
  Best Local Similarity  70.6%;  Pred. No. 5.1e-16;
  Matches  207;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        259 GTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGG 318
              ||  || |||||||| ||| |  |||||||| |    |   |   |  ||||||||
Db      46051 GTTTGCTTGCTTGCTTGCTTGTTTGCTTGCTTACTTGTTTGCTTGTTTGCTTGCTTGCTT 45992

Qy        319 GACAGTAGGGCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTT 378
              |   |   || |||||||| |||  | ||||||||         |   |  |||| ||||
Db      45991 GCTTGCTTGCTTGCTTGCTTGCTCACTTGCTTGCTCGCTCCTTTGCTCGCTTGCTTGCTT 45932

Qy        379 GCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 438
              ||| ||  |  |||| ||| || |||||||||||| ||| || |||||||||||| |||
Db      45931 GCTTGTTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 45872

Qy        439 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 498
              || || |||||||| ||| || || |||||||| | | || || ||||||| | |||
Db      45871 GCTTGTTTGCTTGCTTGCTTGCTTGTTTGCTTGCTTGTTTGCTTGTTTGCTTGTTTGCTT 45812

Qy        499 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGA 551
              || |||||||||||| ||| || |||||||||||| ||| || ||||| || |
Db      45811 GCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTCCTCA 45759


RESULT 73
AL772377
LOCUS       AL772377               217837 bp    DNA     linear   ROD 08-OCT-2002
DEFINITION  Mouse DNA sequence from clone RP23-141B15 on chromosome 2, complete
            sequence.
ACCESSION   AL772377
VERSION     AL772377.6  GI:23893026
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 217837)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 12, 2002 this sequence version replaced gi:23395436.
            ---------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-141B15 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .217837
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
```

```
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-141B15"
                     /clone_lib="RPCI-23"
BASE COUNT    65927 a  46456 c  45477 g  59977 t
ORIGIN

  Query Match             15.5%;  Score 155.4;  DB 10;  Length 217837;
  Best Local Similarity   70.6%;  Pred. No. 5e-16;
  Matches  207;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        259 GTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGG 318
              ||  || |||||||| ||| |  |||||||| |   |   |   |  ||||||||
Db      37219 GTTTGCTTGCTTGCTTGCTTGTTTGCTTGCTTACTTGTTTGCTTGTTTGCTTGCTTGCTT 37278

Qy        319 GACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTT 378
              |   |   || |||||||| |||  | ||||||||          |   |  |||| ||||
Db      37279 GCTTGCTTGCTTGCTTGCTTGCTCACTTGCTTGCTCGCTCCTTTGCTCGCTTGCTTGCTT 37338

Qy        379 GCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 438
              ||| ||  |  |||| ||| || ||||||||||| ||| || |||||||||||| |||
Db      37339 GCTTGTTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 37398

Qy        439 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 498
              || || ||||||||| ||| || || |||||||| | | || || ||||||| | |||
Db      37399 GCTTGTTTGCTTGCTTGCTTGCTTGTTTGCTTGCTTGTTTGCTTGTTTGCTTGTTTGCTT 37458

Qy        499 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGA 551
              || |||||||||||| ||| || |||||||||||| ||| || ||||| || |
Db      37459 GCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTCCTCA 37511


RESULT 74
AC128183
LOCUS       AC128183               196519 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-381B8, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC128183
VERSION     AC128183.3  GI:25081459
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 196519)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196519)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 196519)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:22855982.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GZNP
                Center clone name: CH230-381B8
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 150327 bases at least Q40
                Consensus quality: 152783 bases at least Q30
                Consensus quality: 154221 bases at least Q20
                Estimated insert size: 156901; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
```

```
          * as soon as it is available and the accession number will
          * be preserved.
          *        1   192049: contig of 192049 bp in length
          *   192050   192149: gap of unknown length
          *   192150   196519: contig of 4370 bp in length.
FEATURES             Location/Qualifiers
     source          1. .196519
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-381B8"
     misc_feature    1. .1529
                     /note="wgs_contig"
     misc_feature    1996. .3760
                     /note="wgs_contig"
     misc_feature    6562. .8382
                     /note="wgs_contig"
     misc_feature    188709. .192049
                     /note="wgs_contig"
BASE COUNT    45700 a  35243 c  33203 g  41414 t  40959 others
ORIGIN

  Query Match             15.5%;  Score 155.2;  DB 2;  Length 196519;
  Best Local Similarity   54.3%;  Pred. No. 5.5e-16;
  Matches  313;  Conservative    0;  Mismatches  263;  Indels    0;  Gaps    0;

Qy        299 GTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 358
              ||  |  || ||| |  |  |  .|   |  || ||| | |   |  || ||| |  |
Db      89836 GTTTGTTTGTTTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGTT 89895

Qy        359 GACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 418
              | | |   || || | | ||| | |   |  || | |   |  ||  || |||     |
Db      89896 GGCTGGTTGGTTGGTTGGTTGGTTGGTTGATTGGTTGTCTGGTTGTCTGGTTGTCTAGTT 89955

Qy        419 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 478
              | ||| ||| |||   | | | ||| |||  || | |   |  ||  || |||   | |
Db      89956 GTCTGGTTGGTTGGCTGGTTGTCTGGTTGGCTGGTTGTCTGGTTGGCTGGTTGTCTGGTT 90015

Qy        479 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 538
              | ||| ||| ||| | | | | ||| ||| ||| |   | | | | ||| ||| | | |
Db      90016 GGCTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTTTGTTGGCCGGTTGGTTGGTTGGTT 90075

Qy        539 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCT 598
              | ||| ||| |  |  | | |   || || ||| |||     |  |   |   |  || ||
Db      90076 GGCTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGCT 90135

Qy        599 GATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 658
              | | | ||| | |   |||||   || || | |   || ||   || |||| |   || |
Db      90136 GGTTGGTTGGTTGGTTGGCTGGCTGGTTGGTTGGTTGGTTGGCTGGTTGCTTGGCTGGTT 90195

Qy        659 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCT 718
              | | || |  | |   || || | || || | |   || |  | ||||| | |   |||
Db      90196 GGTTGGTTTGTTGGCTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGCT 90255

Qy        719 GCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 778
              | | |||||  | |   || |  | | ||| | |   || || | ||||| | |   ||||
Db      90256 GGTTGGCTGGTTGGTTGGTTGGTTGACTGGTTGGTTGGTTGGTTGGCTGGTTGCTTGGCT 90315

Qy        779 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTT 838
              | | || || | |   |||||   || || | |   |||||  |||   | ||| ||| ||
Db      90316 GGTTGGTTGGTTGGTTGGCTGGCTGGTTGGTTGGTTGGCTGGCTGGCTGGTTGGTTGGTT 90375

Qy        839 GCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCT 874
              | ||| | |   ||  | |||  || | |  ||||
Db      90376 GGTTGGTTGGTTGGTTGGTTGGTTTTGAGGCT 90411


RESULT 75
AC122009
LOCUS       AC122009              152054 bp    DNA     linear   ROD 16-JUL-2002
DEFINITION  Mus musculus clone RP24-347B22, complete sequence.
ACCESSION   AC122009
VERSION     AC122009.2  GI:21844644
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 152054)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 152054)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 152054)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 152054)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Jul 16, 2002 this sequence version replaced gi:21040133.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information -----------------
            Center project name: M_BB0347B22
            ------------------------------------------------------
FEATURES             Location/Qualifiers
     source          1. .152054
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-347B22"
BASE COUNT    44946 a  33300 c  31206 g  42602 t
ORIGIN

  Query Match             15.5%;  Score 155;  DB 10;  Length 152054;
  Best Local Similarity   63.6%;  Pred. No. 6.4e-16;
  Matches  236;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy        213 GGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGC 272
              || ||   |||| ||| |||| ||| || ||||||||      |   |   || |||||||
Db     126484 GGTTAACTCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 126543

Qy        273 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGC 332
              | ||| || ||| |   |  |   ||     | |||||||      | | |   || |||
Db     126544 TTGCTTGCTTGCCTATGGGTGTTCCATGGTTCTTGCTTGCCTGCTTGCCTGCTTGCTTGC 126603

Qy        333 TTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGC 392
              |||||  ||| || ||||||||      |   |   |  |||| ||||||| |       |
Db     126604 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTATGGGT 126663

Qy        393 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 452
                |      | |||||||||  ||| ||||||||||||||| ||| || ||||||||||
Db     126664 GTTCCATGGTTCTTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 126723

Qy        453 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 512
              | ||| || |||||||||||| ||| || ||||||||||     ||  | |     | |
Db     126724 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTATGGGTGTTCCATGGTTC 126783

Qy        513 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGC 572
              | ||| |||||||||| |||| || || ||||||||      |   |   |  |||||||
Db     126784 TTGCTTGCCTGCTTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 126843

Qy        573 TTGCTAAGGGG 583
              ||||    |||
```

Db 126844 TTGCCTGTGGG 126854

RESULT 76
AC113245
LOCUS AC113245 210964 bp DNA linear HTG 06-MAR-2003
DEFINITION Mus musculus clone RP23-175D5, WORKING DRAFT SEQUENCE, 16 unordered pieces.
ACCESSION AC113245
VERSION AC113245.5 GI:28867081
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 210964)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP23-175D5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 210964)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (27-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 210964)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 6, 2003 this sequence version replaced gi:28629993.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L23486
Center clone name: 175_D_5
---------- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 206691 bases at least Q40
Consensus quality: 208198 bases at least Q30
Consensus quality: 208925 bases at least Q20
Insert size: 220000; agarose-fp
Insert size: 209464; sum-of-contigs
Quality coverage: 8.3 in Q20 bases; agarose-fp
Quality coverage: 8.7 in Q20 bases; sum-of-contigs
----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 622: contig of 622 bp in length
* 623 722: gap of 100 bp
* 723 1420: contig of 698 bp in length
* 1421 1520: gap of 100 bp
* 1521 2183: contig of 663 bp in length
* 2184 2283: gap of 100 bp
* 2284 2935: contig of 652 bp in length
* 2936 3035: gap of 100 bp
* 3036 4527: contig of 1492 bp in length
* 4528 4627: gap of 100 bp
* 4628 8428: contig of 3801 bp in length
* 8429 8528: gap of 100 bp
* 8529 12863: contig of 4335 bp in length
* 12864 12963: gap of 100 bp
* 12964 20933: contig of 7970 bp in length
* 20934 21033: gap of 100 bp
* 21034 34850: contig of 13817 bp in length
* 34851 34950: gap of 100 bp
* 34951 49577: contig of 14627 bp in length
* 49578 49677: gap of 100 bp
* 49678 65321: contig of 15644 bp in length
* 65322 65421: gap of 100 bp
* 65422 84391: contig of 18970 bp in length
* 84392 84491: gap of 100 bp
* 84492 111661: contig of 27170 bp in length
* 111662 111761: gap of 100 bp
* 111762 139035: contig of 27274 bp in length
* 139036 139135: gap of 100 bp
* 139136 178376: contig of 39241 bp in length
* 178377 178476: gap of 100 bp
* 178477 210964: contig of 32488 bp in length.
FEATURES Location/Qualifiers
source 1. .210964
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-175D5"
/clone_lib="RPCI-23 Female Mouse BAC"
misc_feature 1. .622
/note="assembly_fragment"
misc_feature 723. .1420
/note="assembly_fragment"
misc_feature 1521. .2183
/note="assembly_fragment"
misc_feature 2284. .2935

```
                     /note="assembly_fragment"
     misc_feature    3036. .4527
                     /note="assembly_fragment"
     misc_feature    4628. .8428
                     /note="assembly_fragment"
     misc_feature    8529. .12863
                     /note="assembly_fragment"
     misc_feature    12964. .20933
                     /note="assembly_fragment"
     misc_feature    21034. .34850
                     /note="assembly_fragment"
     misc_feature    34951. .49577
                     /note="assembly_fragment"
     misc_feature    49678. .65321
                     /note="assembly_fragment"
     misc_feature    65422. .84391
                     /note="assembly_fragment"
     misc_feature    84492. .111661
                     /note="assembly_fragment"
     misc_feature    111762. .139035
                     /note="assembly_fragment"
     misc_feature    139136. .178376
                     /note="assembly_fragment"
     misc_feature    178477. .210964
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    60649 a  44731 c  44401 g  59683 t   1500 others
ORIGIN

  Query Match             15.4%;  Score 154.2;  DB 2;  Length 210964;
  Best Local Similarity   47.1%;  Pred. No. 7.8e-16;
  Matches  288;  Conservative    0;  Mismatches  323;  Indels    0;  Gaps    0;

Qy        347 GCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTT 406
              | |||||| |   |   | |     | | ||| | | |  ||| |||| | | ||  | ||
Db       7924 GGTTGCTGGGTTGCTGGGATACTGGGTTGCTGGGTTGCTGGGTTGCTGGGTTGCTGGGTT 7983

Qy        407 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 466
              ||| | | ||| |  |||| |  | || | ||||  | | ||| | | |||||  | ||
Db       7984 GCTGGGTTGCTAGGTTGCTGGGATACTGGGTGGCTGGGTGGCTGGGTTGCTGGGATACTG 8043

Qy        467 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 526
              | ||||| | ||||  | | ||| | | |||||   ||| | | ||| | ||||  | ||
Db       8044 GGTTGCTGGGTGGCTGGGTGGCTGGGTGGCTGGGTGGCTGGGTGGCTGGATGGCTGGGTT 8103

Qy        527 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACG 586
              ||| | | |||||  |||| | | |  ||   |  ||| |||| | |||||  |     |
Db       8104 GCTGGGTTGCTGGGTTGCTGGGTTACTGGGTTGCTGGGTTGCTAGGTTGCTGGGATACTG 8163

Qy        587 GTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 646
              |   ||  |  || ||| ||| | ||  |   | |    | | |  ||| ||||||| |
Db       8164 GGTGGCTGGGTTGCTGGGTTGCTGGTTTGCTGGGTTACTGGGTTGCTGGGTTGCTGGGTT 8223

Qy        647 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCT 706
              |||||   |   |       | | |  ||| ||||||| |||| |         | |
Db       8224 GCTAGGTTGCTGGGATACTGGGTGGCTGGGTTGCTGGGTTGCTGGGTTACTGGGTTGCTG 8283

Qy        707 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCT 766
              | | |  ||| | ||||| |||| |   |     | |   | | |  ||| | |||||
Db       8284 GGTTGCTGGGTTACTGGGTTGCTGGGTTGCTGGGTTGCTAGGTTGCTGGGATACTGGGTG 8343

Qy        767 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTG 826
              ||| |  ||   | |||   | | |  ||| | ||||| |||| |  ||   | | ||||
Db       8344 GCTGGGTGGCTGGGTGGCTGGGTTGCTGGGATACTGGGTTGCTGGGTGGCTGGGCGGCTG 8403

Qy        827 GCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTG 886
              | | ||| | | ||| | | |  ||
Db       8404 GATGGCTGGATGGCTGGGTGGCTGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 8463

Qy        887 GGAGAACAGAGTAGGGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGG 946

Db       8464 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 8523

Qy        947 ACACGGGACCC 957
                    | | |
Db       8524 NNNNNGCTCAC 8534


RESULT 77
AL606528
LOCUS       AL606528              148066 bp    DNA     linear   ROD 29-JUN-2002
DEFINITION  Mouse DNA sequence from clone RP23-45H23 on chromosome 13, complete
            sequence.
ACCESSION   AL606528
VERSION     AL606528.7  GI:18873507
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 148066)
  AUTHORS   Oliver,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 22, 2002 this sequence version replaced gi:18476769.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-45H23 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: UK Medical Research Council
            Center code: UK-MRC
            Web site: http://mrcseq.har.mrc.ac.uk
            Contact: mouseq@har.mrc.ac.uk
            --------------.
FEATURES             Location/Qualifiers
     source          1. .148066
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="13"
                     /clone="RP23-45H23"
                     /clone_lib="RPCI-23"
BASE COUNT    43903 a  33152 c  32056 g  38955 t
ORIGIN

  Query Match             15.3%;  Score 153.4;  DB 10;  Length 148066;
  Best Local Similarity   65.1%;  Pred. No. 1.1e-15;
  Matches  226;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy        191 GTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTG 250
              |||| |  ||  |||  |  |  |||| | | ||  |  | |||||
Db      52569 GTGTCACCTCTCCTGAATATGCTTGCCAGGCTTTCAGGCTTAGATGGTTCAGGTTTGCTT 52628

Qy        251 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 310
```

```
                |   |   |||||| || ||||| || ||||||||     | | ||  || |||||
Db      52629 GCTTGCTTGCTTGCCTGCCTGTTCGCTTGCTTGCTTGCTTGCCTGCCTGTTCGCTTGCTT 52688

Qy        311 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 370
              |||     | | ||  |  ||||||| ||| || ||||||||     | | ||  |  |
Db      52689 GCTTGCTTGCCTGTTCGTTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGTTCGTTT 52748

Qy        371 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 430
              ||| |||||||  |   |  |||| ||| || ||||||||||| ||| || ||| || |
Db      52749 GCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGTTC 52808

Qy        431 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              | | ||| || |||||||||||| ||| || |||||||||||| ||| || |||||||||
Db      52809 GTTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 52868

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 537
              ||  | | |  |||||||||||| ||| ||||| | | ||||| |||
Db      52869 GCCTGTTCGTTTGCTTGCTTGCTTGCTTGCCTGTTCGTTTGCTTGCT 52915
```

RESULT 78
AC118179/c
LOCUS AC118179 217604 bp DNA linear HTG 20-NOV-2002
DEFINITION Rattus norvegicus clone CH230-448L22, WORKING DRAFT SEQUENCE, 3 unordered pieces.
ACCESSION AC118179
VERSION AC118179.6 GI:25137770
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 217604)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 217604)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (14-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 217604)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 20, 2002 this sequence version replaced gi:23908319.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GVMT
Center clone name: CH230-448L22
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 170842 bases at least Q40
Consensus quality: 172543 bases at least Q30
Consensus quality: 173598 bases at least Q20
Estimated insert size: 173300; sum-of-contigs estimation
Quality coverage: 10x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 78593: contig of 78593 bp in length
* 78594 78693: gap of unknown length
* 78694 216223: contig of 137530 bp in length
* 216224 216323: gap of unknown length
* 216324 217604: contig of 1281 bp in length.
FEATURES Location/Qualifiers
source 1. .217604
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"

```
                     /clone="CH230-448L22"
     misc_feature    1. .1116
                     /note="wgs_contig"
     misc_feature    1. .814
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BZ253397"
     misc_feature    2909. .3666
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BZ253398"
BASE COUNT    48982 a  39593 c  39097 g  46956 t  42976 others
ORIGIN

  Query Match             15.3%;  Score 153.2;  DB 2;  Length 217604;
  Best Local Similarity   51.8%;  Pred. No. 1.1e-15;
  Matches  347;  Conservative    0;  Mismatches  323;  Indels    0;  Gaps    0;

Qy        219 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 278
              || ||| ||  || | | | || || ||| |  |      |   |  ||  || | | |
Db     188511 GGTTGCTTGGTTGGTAGGTTGCTTGGTTGGTAGGTTACTTGGTTGGTTGACTGGTTGGTT 188452

Qy        279 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 338
              |  || | | |  |  | | |    |  || ||| | ||       |       || ||| |
Db     188451 GTTTGGTAGATTGGCTGGCTGGTTGGTTGTTTGGTAAGTTAGTTGGTAAGTTGGTTGGTA 188392

Qy        339 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 398
              | | |  || | | |  |  |  ||   || || | | ||| | |   || || | | |
Db     188391 GGTTGGTTGGTAGGTTGGTTGGTAGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTA 188332

Qy        399 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 458
              |  || ||| | | | | | |  || ||| | | | | |||  || ||| |   | | |
Db     188331 GGTTGGTTGGTAGGTTGGTTGGTTGTTTGGTAGATTGGTGGGTTGGTTGATAAGTAGGTT 188272

Qy        459 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 518
              |  || ||| ||| | | | |  ||  || |||   | | |  || ||  ||| |   |
Db     188271 GGTTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGCTGGTTGGTTGGTTCTTTGGTAAGTT 188212

Qy        519 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTA 578
              |  || |   |||   | | |  || ||| | |   |   |   || || ||| ||| ||
Db     188211 GGTTGGTAAGTTGGGTGGTAGGTTGGTTGGTTATTTGGTTGGTAGGTTGTTTGGTTGGTA 188152

Qy        579 AGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 638
              | |   |   |  ||  || | | ||| | |   || || | || || | ||  || |
Db     188151 GGTTGGTAGGTTGGTTGGTTGGTAGGTTGGTTGGTTGGTTGGTAGGTTGGTTGTTTGGTT 188092

Qy        639 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 698
              | | || || |||   || || | ||||| | |   ||||| | ||||| | |   || |
Db     188091 GGTAGGTTGGTAGGTTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGCTGGTTGGTTGGTT 188032

Qy        699 CCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCT 758
                |   || ||   || || | || ||   |  || |  | || || | |   ||
Db     188031 GGTTCTTTGGTAAGTTGGTTGGTAGGTTGGGTGGTAGGTTGATTGGTTGTTTGGTTGGTA 187972

Qy        759 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 818
              | | || || |||   || || | || || |||   ||  | | || || ||    || |
Db     187971 GGTTGGTTGGTAGGTTGGTTGGTTGGTTGGTAGATTGGTGGGTTGGTTGATAAGTAGGTT 187912

Qy        819 GCCTGCTGGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTG 878
              |  || | | | | ||| ||| ||| | |   ||  | | |  || ||    || || |
Db     187911 GGTTGGTTGGTAGATTGGTTGGTTGGTTGATAGGTTGATAGGTTGGTAGGTTGGTTGGTT 187852

Qy        879 TGCAGCTGGG 888
                | | | ||
Db     187851 GGGTGGTTGG 187842


RESULT 79
AC106600/c
LOCUS       AC106600               238112 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-227L10, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
ACCESSION   AC106600
VERSION     AC106600.4  GI:30579745
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 238112)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 238112)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 238112)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
```

```
COMMENT     On May 13, 2003 this sequence version replaced gi:23606102.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GLLO
                Center clone name: CH230-227L10
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 225209 bases at least Q40
                Consensus quality: 228391 bases at least Q30
                Consensus quality: 230534 bases at least Q20
                Estimated insert size: 239097; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   233995: contig of 233995 bp in length
            *  233996   234095: gap of unknown length
            *  234096   235636: contig of 1541 bp in length
            *  235637   235736: gap of unknown length
            *  235737   238112: contig of 2376 bp in length.
FEATURES             Location/Qualifiers
     source          1. .238112
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-227L10"
     misc_feature    1. .1415
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    5157. .8726
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    9883. .10719
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BZ105821"
     misc_feature    complement(231646. .232487)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BZ105821"
BASE COUNT    62132 a  48822 c  51700 g  69628 t   5830 others
ORIGIN

  Query Match            15.3%;  Score 152.8;  DB 2;  Length 238112;
  Best Local Similarity  52.9%;  Pred. No. 1.3e-15;
  Matches  351;  Conservative    0;  Mismatches  312;  Indels    1;  Gaps    1;

Qy        215 GTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 274
              ||  || ||  ||  || | | | |  || ||| |  |  |   |  |  || ||| |
Db     133221 GTTTGGTTGGTTGGTTGGTTGATTGGTTGGTTGGTTGGTTGGTTGATTGGTTGGTTGGTT 133162

Qy        275 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 334
              | | |  || ||| |  |  |   |  | ||| ||| |  |  |   |  |  || ||
Db     133161 GGTTGATTGATTGGTTGGTTGGTTGGTAGGCTGGTTGGTTGGTTGGTTGGTTGTTTGGTT 133102

Qy        335 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTC 394
              | | | | |  || ||| |  |  |   |  || || | | ||| | |  || || |
Db     133101 GGTTGGTTGTTTGGTTGGTTGGTTGTTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 133042

Qy        395 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 454
              | |    || ||| ||| | | | |  || ||| ||| | | | |  || ||| ||| |
Db     133041 GGTTTGTTGGTTGTTTGTTTGTTTGGTTGGTTGTTTGGTTGTTTGGTTGGTTGATTGGTT 132982

Qy        455 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 514
              | | |  || ||| ||| | | | |  || ||| ||| |   | |  || ||| ||| |
Db     132981 GTTTGTTTGTTTGTTTGGTTGGTTGGTTGGTTGGTTGGTTAGTTGTTTGGTTGTTTGGTT 132922

Qy        515 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGG-GGACAGTAGGGCTGCTTGCT 573
              | | |  || ||| ||| | | | |  || ||| |  |  ||   ||  |  || ||| |
Db     132921 GGTTGGTTGGTTGGTTGATTGGTTGTTTGTTTGTTTGGTTGGTTTGTTTGTTTGTTTGTT 132862

Qy        574 TGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTA 633
              || |  | |   |   |  |  || || | | ||| | |   || || | || || | |
Db     132861 TGTTTGGTTGGTTGGTTGGTTGGTTGTTTGGTTGGTTGGTTGATTGGTTGGTTGGT 132802

Qy        634 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 693
                | || | || || | |   || || | || || | |   || || | || || | |
Db     132801 TAGTTGTTTGGTTGGTTGGTTGGTTGATTGGTTGGTTGGTTGGTTGGTTGGTTGATTGGT 132742

Qy        694 GGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTA 753
              || |  | || || | |   || || | || || | ||   | |  | || || | |
Db     132741 TGGTTGGTAGGTTGGTTGGTTGGTTGGTTGGTTGGTTGTTTAGTTGGTTGGTTGTTTGGT 132682

Qy        754 GGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTA 813
              || || | |  || | |   || || | || || | |   | |  | || || | ||
Db     132681 TGGTTGGTTGTTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGTTTAGTTGGTTGGTTGTT 132622

Qy        814 GGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGC 873
              || ||  || | | ||| ||| ||| ||| | ||  ||  | |||  || |     ||
Db     132621 TGGTTGGTTGGTTGTTTGGTTGGTTGATTGGTTGTTTGGTTGGTTGGTTGGTTGATTGGT 132562

Qy        874 TGCT 877
              || |
Db     132561 TGTT 132558


RESULT 80
AC120242/c
LOCUS       AC120242              201016 bp    DNA     linear   HTG 12-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-379J13, WORKING DRAFT SEQUENCE.
ACCESSION   AC120242
VERSION     AC120242.3  GI:23908271
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 201016)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
```

Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 201016)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (05-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 201016)

AUTHORS Rat Genome Sequencing Consortium.

TITLE Direct Submission

JOURNAL Submitted (12-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Oct 12, 2002 this sequence version replaced gi:21747233.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

-------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
Center project name: GVTN
Center clone name: CH230-379J13

-------------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 186506 bases at least Q40
Consensus quality: 187889 bases at least Q30
Consensus quality: 188707 bases at least Q20
Estimated insert size: 191116; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 201016: contig of 201016 bp in length.

FEATURES Location/Qualifiers

source 1. .201016
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-379J13"

misc_feature 1. .1118
/note="wgs_end_extension
clone_end:Sp6"

misc_feature 7768. .8481
/note="clone_boundary
clone_end:Sp6
site:MboI
end_sequence:RXAXC55TV"

misc_feature complement(136377. .200482)
/note="clone_boundary
clone_end:T7
site:MboI
end_sequence:RXAXC55TJ"

BASE COUNT 51464 a 48878 c 46169 g 42732 t 11773 others

ORIGIN

Query Match 15.2%; Score 152.6; DB 2; Length 201016;
Best Local Similarity 51.5%; Pred. No. 1.4e-15;
Matches 350; Conservative 0; Mismatches 329; Indels 0; Gaps 0;

```
Qy      198 TGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGAC 257
            |||| ||| |    || |   |  ||| | |||| ||  |   |  |||| | |
Db   178475 TGCTGGTGTGGTGGCGTGGTGGTGTGCTGGTGTGCTGGCGTGGTGGTGCGCTGGTGTGCT 178416

Qy      258 AGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGG 317
            ||  ||  |  |||| |   |   || |||||  | |   ||  |   || |||||  |
Db   178415 GGTGTGCTGGTGTGCTGGTGTGGTGGCGTGCTGGCGTGGTGGTGTGGTGGCGTGCTGGCG 178356

Qy      318 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCT 377
            |   ||  ||  |  ||||||   ||  |  || ||  | |   |   ||      |||
Db   178355 TGGTGGTGTGCTGGTGTGCTCGTGTGCTGGTGTGGTGGTGTGGTGGCGTGGTGCTGTGCT 178296

Qy      378 TGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 437
            ||  |  ||  |||| |   ||  |  |  | |   | |||| || | |  || | |
Db   178295 GGCGTGGTGGTGTGCTGGTGTGCTGGTGTCGTGGTGTGGTGGCGTGGTGGTGTGGTAGTG 178236

Qy      438 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 497
            ||  || || | ||  |||||| || | |  |||| ||  |   |  |||| ||  | |
Db   178235 TGCTGGCGTGGTGGCGTGCTGGCGTGGTGGTGTGCTGGCGTGGTGGTGTGCTGGCGTGGT 178176

Qy      498 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGAC 557
            ||  |||| || || | ||  ||  || || | ||  | |||| || | ||     ||
Db   178175 GGTGTGCTGGCGTGGTGGCGTGCTGGCGTGGTGGCGTGGTGGCGTGGTGGCGTGCTGGCG 178116

Qy      558 AGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGC 617
            |  ||  |||| || ||||   | |  ||   ||  || || |||  || | |   ||
Db   178115 TGGTGGTGTGCTGGCGTGCTGGCGTGGTGGCGTGCTGGCGTGGTGGTGTGCTGGCGTGGT 178056
```

```
Qy        618 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 677
              |   ||  ||  |  ||  ||||||  || | ||  |   |   ||  || .|  ||
Db     178055 GGTGTGGTGGCGTGGTGGTGTGCTGGTGTGGTGGTGTGCTGGTGTGGTGGCATGGTGGTG 177996

Qy        678 TGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 737
              ||||||  || | | | ||   |  ||  ||  |  ||  ||||||  || | |   |
Db     177995 TGCTGGTGTGGTGGCATGGTGGCGTGGTGGCGTGGTGGCGTGCTGGCGTGGTGGCGTGCT 177936

Qy        738 TCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 797
                |  ||  ||  |  ||  || |||  |||| |   ||  |   |   ||  |  ||
Db     177935 GGCGTGGTGGCGTGCTGGCGTGGTGGCGTGCTGGCGTGGTGGTGTGCTGGCGTGGTGGTG 177876

Qy        798 TGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 857
              ||||||  || | ||  |   || || |||  |||| || || | ||  |  ||   | |
Db     177875 TGCTGGCGTGGTGGTGTGCTGGCGTGGTGGTGTGCTGGCGTGGTGGCGTGCTGGCGTGGT 177816

Qy        858 TGCCTGCTACTAGGGCTGC 876
              || || |  |  ||  ||
Db     177815 GGCGTGGTGGTGTGGTGGC 177797
```

RESULT 81
AL669980/c
LOCUS AL669980 157624 bp DNA linear ROD 24-OCT-2002
DEFINITION Mouse DNA sequence from clone RP23-218C2 on chromosome 4, complete sequence.
ACCESSION AL669980
VERSION AL669980.9 GI:24394934
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 157624)
AUTHORS Matthews,L.
TITLE Direct Submission
JOURNAL Submitted (18-OCT-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Oct 25, 2002 this sequence version replaced gi:23337200.
---------- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-218C2 is from the RPCI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.
FEATURES Location/Qualifiers
source 1. .157624
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-218C2"
/clone_lib="RPCI-23"
BASE COUNT 38491 a 38931 c 39485 g 40717 t
ORIGIN

Query Match 15.2%; Score 152.4; DB 10; Length 157624;
Best Local Similarity 52.0%; Pred. No. 1.6e-15;
Matches 367; Conservative 0; Mismatches 336; Indels 3; Gaps 1;

```
Qy        185 GACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGC 244
              | | | || |  ||  | ||  |      ||   |   ||  ||  ||| ||||||  ||
Db     28696 GCCAGGGTTTTGGTTTTGGTTGGTTGGCTGGCTGGCTGGCTGGCTGGCTGGC 28637

Qy        245 TTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGC 304
              | | ||   ||   || ||   | | | | | |||   | ||| ||   ||   || ||
Db     28636 TGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTTGTTGGTTGGTTCGTTGGT 28577

Qy        305 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 364
                | | | ||   ||  ||| | |  ||| | | | || |  ||| | |||  |    |
Db     28576 TAGTTGGTTGGCTGGTTAGTTGACTGGCTGGTTGGTTGACTGGCTGGTTGATTGAGTGGC 28517

Qy        365 AGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 424
              ||   | ||  | | | || |    | | | ||||  | | | |  || |||||  |
Db     28516 TGGTTGGTTGTTTGGTTGGTTGTTTGGTTGGTTGGCTGGTTGGTTGACTGGCTGGTTAGT 28457

Qy        425 TTGCTTGCTCGCTG---GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 481
              |  || ||| ||||   |  || ||| ||| | | |    || ||| ||| | | | | |
Db     28456 TGACTGGCTGGCTGGTTGGTTGGTTGGTTGGTTAGTTGGTTGATTGGTTGATTGAC 28397

Qy        482 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 541
              ||  || ||| | | | |  || ||| |||   |   | ||  ||| ||| | | | |
Db     28396 TGGCTGGTTGGTTGGTTGGTTGGTTGACTGGCTGGCTAGTTGGTTGGTTGGTTGGT 28337

Qy        542 TGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGAT 601
              || ||| |  |  |   |  ||||  |||  || |  |  |   |    || |||||
Db     28336 TGGTTGGTTGGTTGGTTGGTTGGCTAGTTGGCTGGTTGGCTGGTTGGCTAGTTGACTGAT 28277

Qy        602 GGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCT 661
              |  |||  |  ||||| | || || | |   || ||  |||||  |  |||||  |
Db     28276 TGGATGACTGGCTGGCTGGTTGGTTGGTTGGTTGGTTGGCTGGCTGGCTGGCTGGCTGTT 28217

Qy        662 GGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCT 721
              || |  | |  |||| | || ||   |   || |    || || | |   || || |
Db     28216 TGGTTAGTTGGTTGGCTGGTTGGTTGACTGGCTGGATGGATGGTTGGTTGGTTGGTTGGT 28157

Qy        722 GGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCT 781
              || || | |   || |  | |  || | |   || || | ||||| | |   || ||
Db     28156 TGGTTGGTTGGTTGGTTGGTTGATTGGTTGGTTGGTTGGTTGGCTGGTTGGCTGGTTGGC 28097

Qy        782 GGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCT 841
              | ||   |  |||||  |||||   |  |||||  |  | | ||| ||| ||| |
Db     28096 TAGTTGACTGATTGGCTGGCTGGCTGGCTGGCTGGCTGGTTAGTTGGTTGGTTGATTGGT 28037

Qy        842 TGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGG 887
              || | |   ||| |  || ||| |      | || |  || | | |
Db     28036 TGGTTGGTTGGCTGGCTGGCTGTTTGGTTAGTTGGTTGGCTGGTTG 27991
```

RESULT 82
AC117951/c
LOCUS AC117951 262289 bp DNA linear HTG 22-SEP-2002
DEFINITION Rattus norvegicus clone CH230-68L11, *** SEQUENCING IN PROGRESS ***, 5 unordered pieces.
ACCESSION AC117951
VERSION AC117951.6 GI:23194552
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

```
REFERENCE   1  (bases 1 to 262289)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 262289)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 262289)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 19, 2002 this sequence version replaced gi:21745850.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequening reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/).  As a result, the
            sequence may extend beyond the ends of the clone and there may be
            contigs that consist entirely of whole genome shotgun sequence
            reads.  Both end sequences and whole genome shotgun sequence only
            contigs will be indicated in the feature table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GTOP
                Center clone name: CH230-68L11
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 247058 bases at least Q40
                Consensus quality: 248642 bases at least Q30
                Consensus quality: 249755 bases at least Q20
                Estimated insert size: 277580; sum-of-contigs estimation
                Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   245055: contig of 245055 bp in length
            *    245056   245155: gap of unknown length
            *    245156   257363: contig of 12208 bp in length
            *    257364   257463: gap of unknown length
            *    257464   258494: contig of 1031 bp in length
            *    258495   258594: gap of unknown length
            *    258595   259652: contig of 1058 bp in length
            *    259653   259752: gap of unknown length
            *    259753   262289: contig of 2537 bp in length.
FEATURES             Location/Qualifiers
     source          1. .262289
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-68L11"
     misc_feature    1. .1619
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    4372. .5390
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    8354. .8807
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH353262"
     misc_feature    complement(242010. .242869)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH353234"
BASE COUNT    68309 a  53803 c  55598 g  72800 t  11779 others
ORIGIN

  Query Match           15.2%;  Score 152.4;  DB 2;  Length 262289;
  Best Local Similarity   54.6%;  Pred. No. 1.4e-15;
  Matches  368;  Conservative    0;  Mismatches  301;  Indels    5;  Gaps    3;

Qy        194 TGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGG 253
              ||| ||| ||   | |  |  |      || | ||||||  ||   ||||| ||| |
Db     107825 TGCCTGCCTGCCTGCCTTCCTGCCTTCCTTCTTGCCTGCCTGCCTCCCTGCCTGCCTACC 107766

Qy        254 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCT 313
                | |   |||| | | |  ||  |||||| | |   |    | | |   |||||| | |
Db     107765 TTCCTGCCTGCCTACCTCCCTGCCTGCCTGCCTCCCTACTTGCCTGCCTGCCTGCCTCCC 107706

Qy        314 GAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 373
```

```
Db     107705 TGCCTGCCTCCCTGCCTGCCTGCTGTCT-GCCTGCCTGCCTGCCTGCCTCCCTGCCTGCC 107647

Qy        374 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 433

Db     107646 TGCCTGCCTGTCTGCCTGCCTCCCTCCCTGCCTGCCTGCCTGCCTGCCTCCCTGCCTCCC 107587

Qy        434 CGCTGGCCTGCTTGCTTGCTCGCTG---GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490

Db     107586 TGCCTGCCTCCCTGCCTGCCTGCTGTCTGCCTGCCTGCCTGCCTGCCTCCCTGCCTGCCT 107527

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550

Db     107526 GCCTCCCTGCCTGCCTGCTGTCT-GCCTGCCTGCCTGCCTGCCTGTCTCCCTGCCTGCCT 107468

Qy        551 AGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATA 610

Db     107467 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCTGCCTGCCT 107408

Qy        611 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 670

Db     107407 CCCTGCCTGCCTGCCTCCCTGCCTGCCTGCCTGCCTCCCTGCCTGCCTCCCTGCCTGCCT 107348

Qy        671 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 730

Db     107347 GCCTCCCTGCCTGCCTCCCTGCCTGCCTGCCTACCTCCCTGCCTGCCTGCCTCCCTGCTT 107288

Qy        731 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 790

Db     107287 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCTGCCTGCCTCCCT 107228

Qy        791 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGG 850

Db     107227 GCCTGCCTCCCTGCCTCCCTGCCTCCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 107168

Qy        851 GGCCGCTTGCCTGC 864

Db     107167 GCCTGCCTGCCTGC 107154
```

```
RESULT 83
AC110123
LOCUS       AC110123               223892 bp    DNA     linear   HTG 18-SEP-2002
DEFINITION  Rattus norvegicus clone CH230-319K14, *** SEQUENCING IN PROGRESS.
            ***, 4 unordered pieces.
ACCESSION   AC110123
VERSION     AC110123.4  GI:23101151
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 223892)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 223892)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 223892)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 18, 2002 this sequence version replaced gi:21738195.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequening reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/).  As a result, the
            sequence may extend beyond the ends of the clone and there may be
            contigs that consist entirely of whole genome shotgun sequence
            reads.  Both end sequences and whole genome shotgun sequence only
            contigs will be indicated in the feature table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/.
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GQKX
                Center clone name: CH230-319K14
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 162421 bases at least Q40
                Consensus quality: 167280 bases at least Q30
                Consensus quality: 170311 bases at least Q20
                Estimated insert size: 191280; sum-of-contigs estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
```

```
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1     3627: contig of 3627 bp in length
            *      3628     3727: gap of unknown length
            *      3728   220749: contig of 217022 bp in length
            *    220750   220849: gap of unknown length
            *    220850   222487: contig of 1638 bp in length
            *    222488   222587: gap of unknown length
            *    222588   223892: contig of 1305 bp in length.
FEATURES             Location/Qualifiers
     source          1. .223892
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-319K14"
     misc_feature    3728. .5113
                     /note="wgs_contig"
     misc_feature    48568. .50773
                     /note="wgs_contig"
     misc_feature    219069. .220749
                     /note="wgs_contig"
BASE COUNT    48513 a   38895 c   39128 g   46102 t   51254 others
ORIGIN

  Query Match             15.2%;  Score 151.8;  DB 2;  Length 223892;
  Best Local Similarity   51.2%;  Pred. No. 1.9e-15;
  Matches  354;  Conservative    0;  Mismatches  337;  Indels    0;  Gaps    0;

Qy        198 TGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGAC 257
              || |||  ||       |   || ||   |  || | |   |  || ||| |  |  |||
Db      25873 TGGTTGGTAGGTTGGCTGGTTGGTTGGTAGGTTGGTTGGCTGGTTGGTTGGTTGGTTGAC 25932

Qy        258 AGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGG 317
               |   | ||| ||| | | | | ||| |||    |  |  ||   | ||| |||    |
Db      25933 TGGTTGGCTGGTTGGTAGGTTGGCTGGTTGGCTGGTTGGTAGGTTGGCTGGTTGGCTGGT 25992

Qy        318 GGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCT 377
               |  ||   | ||| ||| | | | |   | |||    |  |   |   || || |\| |
Db      25993 TGGTAGGTTGGCTGGTTGGTTGGTTGGTAGGTTGGCTGGTTGGTTGGTTGGTTGGTAGAT 26052

Qy        378 TGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 437
              || |||   || || |   | |  || ||| ||| | | | |  || || |||| | | |
Db      26053 TGGTAGGTTGGTTGATAAGTAGGTTGGTTGGTTGGTAGATTGGTTGGTTCCTTGATAGGT 26112

Qy        438 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 497
               |   | ||| | | | | | |  || ||| ||| | ||| |  || | | ||| | | |
Db      26113 TGATAGGTTGGTAGGTTGGTTGGTTGATTGTTTGGTTGCTAGGTTGGTAGGTTGGTAGGT 26172

Qy        498 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGAC 557
               | || ||| ||| | | | |  || ||| ||| | | | |  || ||| |   |  |
Db      26173 TGTTTGTTTGGTTGGTTGGTAGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGT 26232

Qy        558 AGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGC 617
              ||   || || ||| |   |  |       |   |  ||  || | | | | | |   ||
Db      26233 AGGTTGGTTGGTTGGTAAGTTGGTTCGTTGGTAGGTTGGTTGGTTGGTAGGTTGGTTGGT 26292

Qy        618 TGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 677
              || |  | || |||   || || | ||    | |   || || | || || | |   |
Db      26293 TGGTACGTTGGTAGGTTGGTTGGTTGGTACGTTGGTAGGTTGGTTGGTTGGTTGGTTGAT 26352

Qy        678 TGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 737
               | | || || |||   || |  | || || |||   || || | |   | | |   ||
Db      26353 AGGTTGGTTGATAGGTTGGTTGGTTGGTTGGTAGGTTGGTTGGTTGATAGGTTGGTAGGT 26412

Qy        738 TCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 797
              |  | || || | |   || || | || || | |   || || | |  || | |   ||
Db      26413 TGGTTGGTTGGTTGGTTGGTTGATAGGTTGGTTGGTTGGTTGGTAGATTGGTTGCTTGGT 26472

Qy        798 TGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 857
                | | || |||| |   ||  |  |||| | ||| | | || |||| | |   | | | |
```

```
Db      26473 AGATTGGTTGCTTGGTTGGTAGGTTGCTTGGTTGGTAGGTTACTTGGTTGGTTGACTGGT 26532

Qy        858 TGCCTGCTACTAGGGCTGCTGTGCAGCTGGG 888
                || || |      |  ||    || | | ||
Db      26533 TGGTTGTTTGGTAGATTGGCTGGCTGGTTGG 26563


RESULT 84
AL833804
LOCUS       AL833804              243716 bp    DNA     linear   HTG 14-MAY-2003
DEFINITION  Mus musculus chromosome 2 clone RP23-20A6, *** SEQUENCING IN
            PROGRESS ***.
ACCESSION   AL833804
VERSION     AL833804.14  GI:30722484
KEYWORDS    HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 243716)
  AUTHORS   Griffiths,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 14, 2003 this sequence version replaced gi:29561863.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to confirm this sequence.  Sequence data
            from the whole genome shotgun alone has only been used where it has
            a phred quality of at least 30.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bM20A6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 243636 bases at least Q40
            Consensus quality: 243658 bases at least Q30
            Consensus quality: 243702 bases at least Q20
            Insert size: 243716; sum-of-contigs
            Insert size: 196320; 26.1% error; agarose-fp
            Quality coverage: 9.00x in Q20 bases; sum-of-contigs Quality
            coverage: 11.17x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *         1   243716: contig of 243716 bp in length.
FEATURES             Location/Qualifiers
     source          1. .243716
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-20A6"
                     /clone_lib="RPCI-23"
     misc_feature    1. .243716
                     /note="assembly_fragment:00160
                     clone_end:SP6
                     vector_side:left
                     clone_end:T7
                     vector_side:right"
BASE COUNT    70331 a   54748 c   52890 g   65747 t
```

ORIGIN

```
  Query Match             15.1%;  Score 151.4;  DB 2;  Length 243716;
  Best Local Similarity   52.3%;  Pred. No. 2.1e-15;
  Matches  335;  Conservative    0;  Mismatches  306;  Indels    0;  Gaps    0;

Qy        231 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 290
              | | | |||   | |   ||   ||   || || | | |   |   ||||  ||| ||||
Db     100559 GTTGGTTGGTTGGTTGATTGGCTGGTTGGTTGGTCGGTTGATTAATTGGCTGGCTGGCTG 100618

Qy        291 AGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTT 350
                ||   |  ||     | |||    ||   |||||   | | | | | |||   | |
Db     100619 GCTGGCTGGCTGGTTGCTTGGCTTGTTGGCTGGTAGGTAGGTTGGTTGGTTGGTTGGTTG 100678

Qy        351 GCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTT 410
              | ||   ||  | | |   |||||||| ||| |  ||   | |   ||||  | | |||
Db     100679 GTTGGTTGGTTAATTGACTGGCTGGCTGGCTGGCCGGCTGGTTGCTTGGCTGGTTGGCTG 100738

Qy        411 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 470
              |   |  ||   |  | | | | |||   | |   | ||| | |||   |   | |
Db     100739 GTAGGTAGGTAGGTAGGTTGGTTGGTTGGTTAGTTGATTGGCTGGTTGGTTGGTCGGTTG 100798

Qy        471 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 530
                | ||||||  ||| ||| ||| ||||||  ||| ||| ||| ||||||  ||| |||
Db     100799 ATTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTG 100858

Qy        531 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTAC 590
              ||| |||||     | ||||   ||   |||||   | |   | | |    ||  |||
Db     100859 GCTGGCTGGTTGCTTGGCTGGTTGGCTGGTAGGTAGGTTGATTGGTTGGTTGGTTGGTTG 100918

Qy        591 GCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 650
              ||  ||  | ||||| | ||||   |   | | |   | ||||  |   | ||   ||||
Db     100919 GCTGGCTGGCTGGCTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTGATTGGCTA 100978

Qy        651 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTA 710
              || ||   |  ||   | || | ||   |||||   ||| || |     |||   |||
Db     100979 GTTGGTTGGTCGGTTGGTTAATTGGCTGGCTGGCTGGCTGGTTGCTTGACTGGTTGGCTG 101038

Qy        711 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 770
              |||||   |  ||   | | || ||     |||   | |  | ||   | |||   |
Db     101039 GTAGGTAGGTAGGTAGGTTGGTTGGTTGGTTAGTTGATTGGCTGGTTGGTTGGTCG 101098

Qy        771 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTT 830
              || |      |||   ||| | |||   |||||   ||| |  ||   ||  |||||||
Db     101099 GTTGATTAATTGGCTGGCTGGCAGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTG 101158

Qy        831 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGG 871
              ||| ||| ||| | |  | ||   ||| ||  |    ||||
Db     101159 GCTGGCTGGCTGGTTGCTTGGCTGGCTGGCTGGTAGGTAGG 101199
```

RESULT 85
AC131433
LOCUS AC131433 283593 bp DNA linear HTG 10-MAY-2003
DEFINITION Rattus norvegicus clone CH230-88E19, *** SEQUENCING IN PROGRESS ***, 7 unordered pieces.
ACCESSION AC131433
VERSION AC131433.4 GI:30522205
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 283593)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 283593)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 283593)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 10, 2003 this sequence version replaced gi:24942388.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

```
            ---------- Project Information
                Center project name: GETX
                Center clone name: CH230-88E19
            ---------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 233657 bases at least Q40
                Consensus quality: 237987 bases at least Q30
                Consensus quality: 241136 bases at least Q20
                Estimated insert size: 250075; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3892: contig of 3892 bp in length
            *     3893     3992: gap of unknown length
            *     3993   273529: contig of 269537 bp in length
            *   273530   273629: gap of unknown length
            *   273630   274710: contig of 1081 bp in length
            *   274711   274810: gap of unknown length
            *   274811   276051: contig of 1241 bp in length
            *   276052   276151: gap of unknown length
            *   276152   277399: contig of 1248 bp in length
            *   277400   277499: gap of unknown length
            *   277500   278685: contig of 1186 bp in length
            *   278686   278785: gap of unknown length
            *   278786   283593: contig of 4808 bp in length.
FEATURES             Location/Qualifiers
     source          1. .283593
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-88E19"
     misc_feature    3993. .5736
                     /note="wgs_contig"
BASE COUNT    75651 a  53392 c  53293 g  60758 t  40499 others
ORIGIN

  Query Match             15.1%;  Score 150.8;  DB 2;  Length 283593;
  Best Local Similarity   54.3%;  Pred. No. 2.5e-15;
  Matches  305;  Conservative    0;  Mismatches  257;  Indels    0;  Gaps    0;

Qy        290 GAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCT 349
              | || || | | |   ||   || || |||  ||| ||   |  |   |  | |  || |
Db     141349 GGGGTGATATTTGCTGTGACAGCAGATGGGGGAGTGGGTAGGTATTTGCTGTTGGTTGGT 141408

Qy        350 TGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCT 409
              || |  |  |   |   || || | | | | |  |||  | | | | |  || |||||
Db     141409 TGGTTGGTTGGTTGGTTGGTTGGTTGGTGGGTGGGTGGGTAGGTGGGTTGGATGGTTGCT 141468

Qy        410 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 469
              |||| ||| || ||||||||||||| ||| || ||||||||||||| ||| || ||||||
Db     141469 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGGA 141528

Qy        470 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 529
              || | |   |  || ||  ||| | | | |  || ||| ||| | | | |  ||  || |
Db     141529 TGGTTGGCTGGTTGGTTAGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGATGGT 141588

Qy        530 TGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTA 589
              || | | |  |  ||| |  |  |   |  || || ||| ||| |     |   |
Db     141589 TGGTTGGTTGGATCGTTGGTTGGTTGGATGGTTGGTTGGTTGGTTGGTTGCATGGTTGGT 141648

Qy        590 CGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 649
              |  ||  | || | ||| | |   || || | || ||   |  ||| || | || || |
Db     141649 TGGTTGGATCATTGGTTGGTTGGTTGGTTGGTTGGATGGTGGGTTGGTTGGTTGGT 141708

Qy        650 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCT 709
              |   || || | || ||   |  ||| || | || ||   |   || |  | || || |
Db     141709 TGGGTGGTTGTTTGGTTGGATGGTGGGTTGTTTGGTTGGATGGTTGGTTGGTTGGATGGT 141768

Qy        710 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 769
              |    | ||   || || | |   || |  | || || | | | || |  | || ||
Db     141769 TGGTTTGTTGGATGGTTGATTGGTTGGATGGTTGGTTGGTTGGATGGTTAGTTGGTTGAA 141828

Qy        770 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCT 829
              |   |  |||| |  |||| |   || | |  | || | |   |  ||  ||   | |
Db     141829 TGGTTGCTTGCTTGCTTGCTTGCTTGGATATTTCGTTGGTTGGTTGATTGGTTGGATGGT 141888

Qy        830 TGCTTGCTTGCTTGCTAGTGGG 851
              || ||| | | |   |||| ||
Db     141889 TGTTTGTTGGATGTTTAGTTGG 141910


RESULT 86
AL645694
LOCUS       AL645694              189236 bp    DNA     linear   HTG 19-JUN-2002
DEFINITION  Mus musculus chromosome 13 clone RP23-193M3, *** SEQUENCING IN
            PROGRESS ***, 6 unordered pieces.
ACCESSION   AL645694
VERSION     AL645694.5  GI:18151524
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 189236)
  AUTHORS   Matthews,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 15, 2002 this sequence version replaced gi:17017921.
            ---------- Genome Center
            Center: UK Medical Research Council
            Center code: UK-MRC
            Web site: http://mrcseq.har.mrc.ac.uk
            Contact: mouseq@har.mrc.ac.uk
            ---------- Project Information
            Center project name: bM193M3
            ---------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 187721 bases at least Q40
            Consensus quality: 188054 bases at least Q30
            Consensus quality: 188428 bases at least Q20
            Insert size: 188736; sum-of-contigs
            Insert size: 192888; 4.5% error; agarose-fp
            Quality coverage: 10.92x in Q20 bases; sum-of-contigs Quality
            coverage: 12.08x in Q20 bases; agarose-fp
            ----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    34757: contig of 34757 bp in length
            *    34758    34857: gap of 100 bp
            *    34858    54711: contig of 19854 bp in length
            *    54712    54811: gap of 100 bp
            *    54812   117926: contig of 63115 bp in length
            *   117927   118026: gap of 100 bp
            *   118027   131915: contig of 13889 bp in length
            *   131916   132015: gap of 100 bp
            *   132016   176088: contig of 44073 bp in length
```

```
            *   176089   176188: gap of 100 bp
            *   176189   189236: contig of 13048 bp in length.
FEATURES             Location/Qualifiers
     source          1. .189236
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="13"
                     /clone="RP23-193M3"
                     /clone_lib="RPCI-23"
     misc_feature    1. .34757
                     /note="assembly_fragment:03383
                     clone_end:SP6
                     vector_side:left"
     misc_feature    34858. .54711
                     /note="assembly_fragment:03309.0"
     misc_feature    54812. .117926
                     /note="assembly_fragment:01236
                     fragment_chain:1"
     misc_feature    118027. .131915
                     /note="assembly_fragment:02240
                     fragment_chain:1"
     misc_feature    132016. .176088
                     /note="assembly_fragment:01751
                     fragment_chain:1"
     misc_feature    176189. .189236
                     /note="assembly_fragment:01679
                     fragment_chain:1
                     clone_end:T7
                     vector_side:right"
BASE COUNT    58348 a  41930 c  40027 g  48429 t    502 others
ORIGIN

  Query Match             15.0%;  Score 150.4;  DB 2;  Length 189236;
  Best Local Similarity   64.8%;  Pred. No. 3.2e-15;
  Matches  223;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy        191 GTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTG 250
              ||||     ||  |||      |  |   ||||  | | ||          |  | |||||
Db     154196 GTGTCACCTCTCCTGAATATGCTTGCCAGGCTTTCAGGCTTAGATGGTTCAGGTTTGCTT 154255

Qy        251 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTT 310
              |    |   |||||| || ||||| || ||||||||     | | ||  || |||||
Db     154256 GCTTGCTTGCTTGCCTGCCTGTTCGCTTGCTTGCTTGCTTGCCTGCCTGTTCGCTTGCTT 154315

Qy        311 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCT 370
              |||      | | ||  |  ||||||| ||| || ||||||||     | | ||  |  |
Db     154316 GCTTGCTTGCCTGTTCGTTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGTTCGTTT 154375

Qy        371 GCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 430
              ||| |||||||  |    |  |||| ||| || ||||||||||| ||| || ||| || |
Db     154376 GCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCTGTTC 154435

Qy        431 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 490
              | | ||| || |||||||||||| ||| || |||||||||||| ||| || |||||||  |
Db     154436 GTTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCCT 154495

Qy        491 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 534
              | ||| | || |||||||||||| ||  |    | ||||||||||
Db     154496 GTTCGTTTGCTTGCTTGCTTGCTTGCCTGTTCGTTTGCTTGCTC 154539


RESULT 87
AC108277_1/c
WPCOMMENT
  Sequence split into 5 fragments  LOCUS AC108277 Accession AC108277
     Fragment Name      Begin     End
     AC108277_0             1    110000
     AC108277_1        100001    210000
     AC108277_2        200001    310000
     AC108277_3        300001    410000
     AC108277_4        400001    492135
Continuation (2 of 5) of AC108277 from base 100001 (AC108277 Rattus norvegicus clone CH2

  Query Match             15.0%;  Score 150.2;  DB 2;  Length 110000;
  Best Local Similarity   64.6%;  Pred. No. 4e-15;
  Matches  224;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy        185 GACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGC 244
              || ||| |    |    |||  |  |  ||    | |||||||| |||  ||  ||||||
Db      69660 GATGGCATTGCTGCCACTGTAGGTTAGTGGTAGTGCCTGCCTGCTTGCCTGCCTGCCTGC 69601

Qy        245 TTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGC 304
              ||||       | | |   ||||||||||  ||| |||||||||       | | |   ||
Db      69600 TTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGC 69541

Qy        305 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 364
              ||||||||       | | |   |||||| |||  ||  || ||| ||||      |   |
Db      69540 CTGCTTGCCTGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGC 69481

Qy        365 AGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 424
                | ||||  ||||||  |   |  |||  ||| ||||||  |||||||  ||| |||| |
Db      69480 CTGCCTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTTC 69421

Qy        425 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 484
              |||  |||   ||  ||||||  |||  |||   ||   ||  |||  |||||||   ||| ||||||
Db      69420 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGC 69361

Qy        485 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 531
              ||||  ||||  ||   ||||||  |||||||   |||  ||||||||||||||
Db      69360 TTGCCTGCTTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCTTG 69314


RESULT 88
AE014176_0/c
WPCOMMENT
  Sequence split into 4 fragments  LOCUS AE014176 Accession AE014176
     Fragment Name      Begin     End
     AE014176_0             1    110000
     AE014176_1        100001    210000
     AE014176_2        200001    310000
     AE014176_3        300001    395229
LOCUS       AE014176               395229 bp    DNA     linear   ROD 06-AUG-2002
DEFINITION  Mus musculus piebald deletion region section 4 of 11 of the
            complete sequence.
ACCESSION   AE014176 AE013600
VERSION     AE014176.1  GI:22128031
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 395229)
  AUTHORS   Peterson,K., King,B., Hagge-Greenberg,A., Roix,J., Bult,C. and
            O'Brien,T.
  TITLE     Functional and comparative genomic analysis of the piebald deletion
            region of mouse chromosome 14
  JOURNAL   Genomics 80 (2), 172 (2002)
  MEDLINE   22150871
   PUBMED   12160731
REFERENCE   2  (bases 1 to 395229)
  AUTHORS   Adams,M. and Mural,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAY-2002) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     The strings of n's in this record represent gaps between contigs,
            and the length of each string corresponds to the length of the gap
            except where there are 20 n's.  A string of 20 n's represents a gap
            of unknown length.
FEATURES             Location/Qualifiers
     source          1. .395229
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="14"
                     /map="Mit164-Mit165"
```

```
     gene            <109204. .>207970
                     /gene="Scel"
     mRNA            <109204. .>207970
                     /gene="Scel"
                     /product="sciellin"
                     /note="partial transcript sequence derived from alignment
                     to Mus musculus NCBI RefSeq sequence NM_022886"
     mRNA            <245271. .>299392
                     /product="TC255155"
                     /note="partial transcript derived from alignment to
                     tentative consensus sequence TC200980, from TIGR Mouse
                     Gene Index (v. 5.0); developmental expression detected by
                     RT-PCR using panel of embryonic RNAs  (E8.5, E10.5, E12.5,
                     E14.5, E16.5 and E18.5) from C57BL/6J mice"
BASE COUNT   101118 a  74785 c  76350 g 109777 t  33199 others
ORIGIN

  Query Match             15.0%;  Score 150.2;  DB 10;  Length 110000;
  Best Local Similarity   53.9%;  Pred. No. 4e-15;
  Matches  355;  Conservative    0;  Mismatches  298;  Indels    6;  Gaps    2;

Qy        206 AGACACCGGGTACGGCTGCCTGCCTG-CTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGC 264
              || |  |  |   | ||||||||||| ||  ||| |   ||  ||||  ||   |  |||
Db      60943 AGCCTGCCTGCCTGCCTGCCTGCCTGCCTGACTGACTGACTGACTGACTGGCTGGCTGGC 60884

Qy        265 CTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGT 324
               ||| |||  ||| |   || |||||  |    |  | |  |||  ||||  |    |
Db      60883 TGGCTGGCTGACTGACTGACTGGCTGACTGACTGGCTGACTGGCTGACTGACTGACTGGC 60824

Qy        325 AGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGT 384
              | |   ||  ||  |||||   ||  ||||  ||   |  ||    ||| || |||
Db      60823 TGACTGACTGACTGACTGGCTGACTGACTGACTGGCTGGCTGGCTGACTGACTGGCTGAC 60764

Qy        385 AGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 444
              ||   |||  |||||  |||  ||  || |||| |   || |||  || ||||||   |
Db      60763 TGGTTGGCTGACTGGCTGGCTGACTGACTGGCTGACTGACTGGCTGACTGGCTGGCTGAC 60704

Qy        445 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 504
              |  || |||  ||| |  |||  ||  ||  ||| |   ||  ||  ||  |||||  ||
Db      60703 TGACTGGCTGACTGACTGGCTGACTGACTGACTGACTGACTGACTGACTGGCTGGCTGGC 60644

Qy        505 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGG 564
              | ||| ||| ||||||  ||| |||  || |||||  ||| ||||  ||   |  ||
Db      60643 TGGCTGGCTGGCTGGCTGGCTGGCTGACTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGC 60584

Qy        565 CTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGG 624
               ||| ||| |||     ||| |   | ||| ||||  |  ||   |   |||||   ||
Db      60583 TGGCTGGCTGGCT-----GACTGACTGACTGACTGACTGACTGGCTGGCTGGCTGGCTGG 60529

Qy        625 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 684
              |||   |  |||||  |||||   |   |||||  |||||   |   | |||   |
Db      60528 CTGGCTGGCTGGCTGACTGGCTGGCTGGCTGGCTGGCTGACTGACTGACTGACTGA 60469

Qy        685 CTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGG 744
              |||   |   | ||   | |||   |   | |||  |||||   |   | ||    |
Db      60468 CTGACTGACTGACTGACTGACTGACTGACTGACTGGCTGGCTGACTGACTGGCTGA 60409

Qy        745 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 804
              ||| | |   |||||  |||||   |   |||||  | |||   |   | |||   |
Db      60408 CTGGTTGGCTGGCTGGCTGGCTGGCTGACTGGCTGACTGACTGACTGACTGACTGACTGA 60349

Qy        805 CTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTG 863
              |||   |   | ||| |||   |  ||  || |||  ||   |   ||| |  || |||
Db      60348 CTGACTGACTGACTGACTGACTGACTGACTGGTTGGCTGGCTGGCTGGCTGACTGACTG 60290


RESULT 89
AC114585/c
LOCUS       AC114585              189576 bp    DNA     linear   HTG 17-MAR-2003
DEFINITION  Mus musculus clone RP23-296J16, WORKING DRAFT SEQUENCE, 10
            unordered pieces.
ACCESSION   AC114585
VERSION     AC114585.7  GI:28975893
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 189576)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-296J16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 189576)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 189576)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 17, 2003 this sequence version replaced gi:28412057.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L23692
                Center clone name: 296_J_16
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
```

```
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 186712 bases at least Q40
            Consensus quality: 187848 bases at least Q30
            Consensus quality: 188381 bases at least Q20
            Insert size: 194000; agarose-fp
            Insert size: 188676; sum-of-contigs
            Quality coverage: 8.0 in Q20 bases; agarose-fp
            Quality coverage: 8.2 in Q20 bases; sum-of-contigs
          ----------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 10 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1    47737: contig of 47737 bp in length
          *   47738    47837: gap of 100 bp
          *   47838    48894: contig of 1057 bp in length
          *   48895    48994: gap of 100 bp
          *   48995    49945: contig of 951 bp in length
          *   49946    50045: gap of 100 bp
          *   50046    51530: contig of 1485 bp in length
          *   51531    51630: gap of 100 bp
          *   51631    57122: contig of 5492 bp in length
          *   57123    57222: gap of 100 bp
          *   57223    66812: contig of 9590 bp in length
          *   66813    66912: gap of 100 bp
          *   66913   136223: contig of 69311 bp in length
          *  136224   136323: gap of 100 bp
          *  136324   146728: contig of 10405 bp in length
          *  146729   146828: gap of 100 bp
          *  146829   185215: contig of 38387 bp in length
          *  185216   185315: gap of 100 bp
          *  185316   189576: contig of 4261 bp in length.
FEATURES             Location/Qualifiers
     source          1. .189576
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-296J16"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .47737
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    47838. .48894
                     /note="assembly_fragment"
     misc_feature    48995. .49945
                     /note="assembly_fragment"
     misc_feature    50046. .51530
                     /note="assembly_fragment"
     misc_feature    51631. .57122
                     /note="assembly_fragment"
     misc_feature    57223. .66812
                     /note="assembly_fragment"
     misc_feature    66913. .136223
                     /note="assembly_fragment"
     misc_feature    136324. .146728
                     /note="assembly_fragment"
     misc_feature    146829. .185215
                     /note="assembly_fragment"
     misc_feature    185316. .189576
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    58607 a  38256 c  38652 g  53160 t    901 others
ORIGIN

  Query Match             15.0%;  Score 150.2;  DB 2;  Length 189576;
  Best Local Similarity   53.7%;  Pred. No. 3.5e-15;

  Matches  354;  Conservative    0;  Mismatches  303;  Indels    2;  Gaps    2;

Qy        232 CTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGA 291
              || |||| |  |||  ||||  |    |  | |   ||  ||  |||||   ||  ||||
Db     172936 CTGGCTGACTGGCTGACTGACTGACTGGCTGACTGACTGACTGACTGGCTGACTGACTGA 172877

Qy        292 GGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTG 351
              ||   |  |||   ||  |||   |    || |||   || ||| |||| |   || |
Db     172876 CTGGCTGGCTGGCTGACTGACTGGCTGACTGGTTGGCTGACTGGCTGGCTGACTGACTGG 172817

Qy        352 CTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTG 411
              ||||  |    |  |    |||||||  ||    ||    ||  |||||   ||  ||
Db     172816 CTGACTGACTGGCTGACTGGCTGGCTGACTGACTGGCTGACTGACTGGCTGACTGACTGA 172757

Qy        412 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 471
              ||  ||| |   ||  ||  || ||||||  ||| ||| ||| ||||||  |||  || |
Db     172756 CTGACTGACTGACTGACTGACTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGACTGACTGG 172697

Qy        472 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 531
              || |||||  ||| ||| ||| ||||||  ||| ||| |||  |||  ||  ||
Db     172696 CTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGACTGACTGACTGACTGA 172637

Qy        532 CTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACG 591
              ||  |||||  ||| |||| |  | | |   |||||  ||  ||    |  | | |   |
Db     172636 CTGACTGGCTGGCTGGCTG-GCTGGCTGGCTGGCTGGCTGACTGGCTGGCTGGCTGGCTG 172578

Qy        592 CCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAG 651
              |||  ||||  |  |||  |   | |||   | |||   |   | |||   | |||   |
Db     172577 GCTGGCTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTG 172518

Qy        652 TAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAG 711
              | |||  |||||   |   |||||  || ||   |   ||||   |||||   |
Db     172517 ACTGACTGGCTGGCTGACTGACTGGCTGACTGGTTGGCTGGCTGGCTGGCTGGCTGACTG 172458

Qy        712 TAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAG 771
              | |||   | |||   |   | ||   | |||   |   | |||   | |||  | |
Db     172457 GCTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGGTTG 172398

Qy        772 TAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTG-CTGGCTT 830
              |||||  |||||  |  |||||  | |||  |  |||||  ||| ||||||
Db     172397 GCTGGCTGGCTGGCTGACTGACTGGCTGACTGACTGACTGACTGGCTGACTGACTGGCTA 172338

Qy        831 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGA 889
              || | | ||| |||  | |  |     | | || || |  | |    ||||  |||
Db     172337 ACTGGTTGGCTGGCTGAGGAGAACCTGCTCATTCTCCTGGCCATACCCGACAGCCAGGA 172279


RESULT 90
AC132758
LOCUS       AC132758              238255 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-101A12, *** SEQUENCING IN PROGRESS
            ***, 4 unordered pieces.
ACCESSION   AC132758
VERSION     AC132758.3  GI:25073372
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 238255)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
```

```
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 238255)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 238255)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:23195993.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KBYE
                Center clone name: CH230-101A12
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 225270 bases at least Q40
                Consensus quality: 227698 bases at least Q30
                Consensus quality: 229450 bases at least Q20
                Estimated insert size: 226046; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    15508: contig of 15508 bp in length
            *    15509    15608: gap of unknown length
            *    15609   224342: contig of 208734 bp in length
            *   224343   224442: gap of unknown length
            *   224443   225652: contig of 1210 bp in length
            *   225653   225752: gap of unknown length
            *   225753   238255: contig of 12503 bp in length.
FEATURES             Location/Qualifiers
     source          1. .238255
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-101A12"
     misc_feature    1. .1480
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    1361. .2115
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BH306826"
     misc_feature    15609. .16703
                     /note="wgs_contig"
     misc_feature    38908. .40937
                     /note="wgs_contig"
     misc_feature    41382. .43075
                     /note="wgs_contig"
     misc_feature    222951. .224342
                     /note="wgs_contig"
BASE COUNT    63779 a  53533 c  52534 g  60451 t   7958 others
ORIGIN

  Query Match             15.0%;  Score 150.2;  DB 2;  Length 238255;
  Best Local Similarity   64.6%;  Pred. No. 3.3e-15;
  Matches  224;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy        185 GACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGC 244
              || ||| |    |    ||| | | ||    | |||||||| |||  ||  ||||||
Db      26656 GATGGCATTGCTGCCACTGTAGGTTAGTGGTAGTGCCTGCCTGCTTGCCTGCCTGCCTGC 26715

Qy        245 TTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGC 304
              ||||      | | |   |||||||||  ||| ||||||||||      | | |   ||
Db      26716 TTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGC 26775

Qy        305 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 364
              ||||||||      | | |   |||||| |||  ||  || ||| ||||     |   |
Db      26776 CTGCTTGCCTGCTTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGC 26835

Qy        365 AGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 424
                | ||||  ||||||  |   |  |||  ||| |||||| |||||||  ||| |||| |
Db      26836 CTGCCTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTTC 26895

Qy        425 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 484
              ||| |||  ||  |||||| ||| |||  ||  || ||| |||||||  ||| ||||||
Db      26896 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGC 26955
```

```
Qy       485 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 531
             |||| |||| ||  |||||| |||||||  ||| ||||||||||||
Db     26956 TTGCCTGCTTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCTTG 27002


RESULT 91
AC116206
LOCUS       AC116206              240200 bp    DNA     linear   HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-10104, WORKING DRAFT SEQUENCE, 5
            unordered pieces.
ACCESSION   AC116206
VERSION     AC116206.4  GI:24818933
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 240200)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 240200)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 240200)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 9, 2002 this sequence version replaced gi:23101726.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GMPU
                Center clone name: CH230-10104
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 227436 bases at least Q40
                Consensus quality: 229834 bases at least Q30
                Consensus quality: 231317 bases at least Q20
                Estimated insert size: 232522; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    62473: contig of 62473 bp in length
            *    62474    62573: gap of unknown length
            *    62574   236709: contig of 174136 bp in length
            *   236710   236809: gap of unknown length
            *   236810   237884: contig of 1075 bp in length
            *   237885   237984: gap of unknown length
            *   237985   239022: contig of 1038 bp in length
            *   239023   239122: gap of unknown length
            *   239123   240200: contig of 1078 bp in length.
FEATURES             Location/Qualifiers
     source          1. .240200
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-10104"
     misc_feature    1. .1428
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    2029. .2702
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH288965"
     misc_feature    complement(53971. .54130)
                     /note="clone_boundary
```

```
                clone_end:Sp6
                site:EcoRI
                end_sequence:BH288967"
BASE COUNT    64992 a  53941 c  53130 g  60171 t   7966 others
ORIGIN

  Query Match             15.0%;  Score 150.2;  DB 2;  Length 240200;
  Best Local Similarity   64.6%;  Pred. No. 3.3e-15;
  Matches  224;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy        185 GACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGC 244
              || ||| |    |    |||  |  |  ||    | |||||||| |||  ||  ||||||
Db      11690 GATGGCATTGCTGCCACTGTAGGTTAGTGGTAGTGCCTGCCTGCTTGCCTGCCTGCCTGC 11749

Qy        245 TTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGC 304
              ||||       | | |   ||||||||||  ||| ||||||||||        | |  |   ||
Db      11750 TTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGC 11809

Qy        305 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 364
              ||||||||      | | |   ||||||  |||  ||  ||  |||  ||||     |   |
Db      11810 CTGCTTGCCTGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGC 11869

Qy        365 AGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 424
                | ||||  ||||||  |   |  |||  ||| ||||||  |||||||  ||| |||| |
Db      11870 CTGCCTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTTC 11929

Qy        425 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 484
              |||  |||   ||  ||||||  |||  |||   ||  ||  |||  |||||||  ||| ||||||
Db      11930 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTGC 11989

Qy        485 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 531
              ||||  ||||  ||  ||||||  |||||||  |||  ||||||||||||
Db      11990 TTGCCTGCTTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCTTG 12036


RESULT 92
AC103570
LOCUS       AC103570              242974 bp    DNA     linear   HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-10M24, WORKING DRAFT SEQUENCE.
ACCESSION   AC103570
VERSION     AC103570.9  GI:24818028
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 242974).
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 242974)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 242974)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 9, 2002 this sequence version replaced gi:22772757.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GDRO
                Center clone name: CH230-10M24
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 231206 bases at least Q40
                Consensus quality: 232201 bases at least Q30
                Consensus quality: 232817 bases at least Q20
                Estimated insert size: 239753; sum-of-contigs estimation
                Quality coverage: 9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
```

```
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *         1   242974: contig of 242974 bp in length.
FEATURES             Location/Qualifiers
     source          1. .242974
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-10M24"
     misc_feature    1. .2423
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    3420. .3992
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH308241"
     misc_feature    238677. .242974
                     /note="wgs_contig"
BASE COUNT    65335 a  52835 c  51454 g  63642 t   9708 others
ORIGIN

  Query Match             15.0%;  Score 150.2;  DB 2;  Length 242974;
  Best Local Similarity   64.6%;  Pred. No. 3.3e-15;
  Matches  224;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy        185 GACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGC 244
              || ||| |    |    ||| |  |  ||    | |||||||| |||  || | ||||||
Db     194611 GATGGCATTGCTGCCACTGTAGGTTAGTGGTAGTGCCTGCCTGCTTGCCTGCCTGCCTGC 194670

Qy        245 TTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGC 304
              ||||      | | |   |||||||||| ||| ||||||||||      | | |   ||
Db     194671 TTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGC 194730

Qy        305 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 364
              |||||||      | | |   |||||| |||  ||  || ||| ||||     |   |
Db     194731 CTGCTTGCCTGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGC 194790

Qy        365 AGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 424
                | ||||  ||||||  |   |  |||  ||| ||||||  |||||||  ||| |||| |
Db     194791 CTGCCTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTTC 194850

Qy        425 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 484
              ||| |||  ||  |||||| ||| |||  ||  || ||| |||||||  ||| ||||||
Db     194851 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCCTGC 194910

Qy        485 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 531
              |||| |||| ||  |||||| |||||||  ||| ||||||||||||
Db     194911 TTGCCTGCTTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCTTG 194957


RESULT 93
AC116072/c
LOCUS       AC116072               244301 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-249G11, *** SEQUENCING IN PROGRESS
            ***, 4 unordered pieces.
ACCESSION   AC116072
VERSION     AC116072.5  GI:25078938
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 244301)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 244301)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 244301)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:22857224.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
```

```
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
          ---------- Project Information
            Center project name: GNEY
            Center clone name: CH230-249G11
          ---------- Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 205700 bases at least Q40
            Consensus quality: 209464 bases at least Q30
            Consensus quality: 211942 bases at least Q20
            Estimated insert size: 222435; sum-of-contigs estimation
            Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
          ----------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 4 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1   239389: contig of 239389 bp in length
          *   239390  239489: gap of unknown length
          *   239490  240661: contig of 1172 bp in length
          *   240662  240761: gap of unknown length
          *   240762  242103: contig of 1342 bp in length
          *   242104  242203: gap of unknown length
          *   242204  244301: contig of 2098 bp in length.
FEATURES             Location/Qualifiers
     source          1. .244301
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-249G11"
     misc_feature    1. .1076
                     /note="wgs_contig"
     misc_feature    3053. .4222
                     /note="wgs_contig"
     misc_feature    83375. .84860
                     /note="wgs_contig"
     misc_feature    238099. .239389
                     /note="wgs_contig"
BASE COUNT    59847 a  48345 c  47032 g  58326 t  30751 others
ORIGIN

  Query Match            15.0%;  Score 150.2;  DB 2;  Length 244301;
  Best Local Similarity  64.6%;  Pred. No. 3.2e-15;
  Matches  224;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy        185 GACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGC 244
              || ||| |    |    ||| |  |  ||    | |||||||| |||  ||  ||||||
Db     132274 GATGGCATTGCTGCCACTGTAGGTTAGTGGTAGTGCCTGCCTGCTTGCCTGCCTGCCTGC 132215

Qy        245 TTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGC 304
              ||||      | | |   |||||||||  ||| ||||||||||      | | |   ||
Db     132214 TTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGCTTGC 132155

Qy        305 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 364
              ||||||||      | | |   ||||||  |||  ||  ||  ||| ||||     |   |
Db     132154 CTGCTTGCCTGCTTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGC 132095

Qy        365 AGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 424
                | ||||  ||||||  |   |  |||  ||| ||||||  |||||||  ||| |||| |
Db     132094 CTGCCTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTGCCTGCTTGCCTGCTTGCCTTC 132035

Qy        425 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 484
              ||| |||  ||  ||||||  ||| |||  ||  ||  || |||  |||||||  ||| ||||||
Db     132034 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCCTGC 131975

Qy        485 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 531
```

```
                  |||| |||| ||  ||||||  |||||||  ||| ||||||||||||
Db     131974 TTGCCTGCTTGCCTGCCTGCCTGCTTGCCTGCTTGCCTGCTTGCTTG 131928


RESULT 94
AC095003
LOCUS       AC095003               238046 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-6K12, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC095003
VERSION     AC095003.7  GI:30467553
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 238046)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 238046)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
```

```
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3  (bases 1 to 238046)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
COMMENT   On May 9, 2003 this sequence version replaced gi:24818271.
          The sequence in this assembly is a combination of BAC based reads
          and whole genome shotgun sequencing reads assembled using Atlas
          (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
          in the feature table below represents a scaffold in the Atlas
          assembly (a 'contig-scaffold'). Within each contig-scaffold,
          individual sequence contigs are ordered and oriented, and separated
          by sized gaps filled with Ns to the estimated size. The sequence
          may extend beyond the ends of the clone and there may be sequence
          contigs within a contig-scaffold that consist entirely of whole
          genome shotgun sequence reads. Both end sequences and whole genome
          shotgun sequence only contigs will be indicated in the feature
          table.
          -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
              Center project name: GBYH
              Center clone name: CH230-6K12
          -------------- Summary Statistics
              Assembly program: Atlas;
              Consensus quality: 226044 bases at least Q40
              Consensus quality: 227924 bases at least Q30
              Consensus quality: 229373 bases at least Q20
              Estimated insert size: 234613; sum-of-contigs estimation
              Quality coverage: 9x in Q20 bases; sum-of-contigs estimation
          ----------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 2 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1   236832: contig of 236832 bp in length
          *   236833   236932: gap of unknown length
          *   236933   238046: contig of 1114 bp in length.
FEATURES             Location/Qualifiers
     source          1. .238046
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-6K12"
     misc_feature    1. .1680
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    6066. .6753
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH360832"
     misc_feature    228068. .228893
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH360837"
     misc_feature    229310. .230352
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    235657. .236832
                     /note="wgs_end_extension
                     clone_end:Sp6"
BASE COUNT    61790 a  49774 c  51497 g  67565 t   7420 others
ORIGIN

  Query Match             15.0%;  Score 149.8;  DB 2;  Length 238046;
  Best Local Similarity   55.7%;  Pred. No. 3.8e-15;
  Matches  306;  Conservative    0;  Mismatches  242;  Indels    1;  Gaps    1;

Qy        299 GTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 358
              || ||   ||| | ||   ||   || |||  | | ||| | ||||  | | |||| | |
Db     167437 GTTGGTTGGCTGGTTGGTTGGCTGGTTGGCTGGTTGGCTGGTTGGCTGGTTGGCTGGTTG 167496

Qy        359 GACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 418
              |   |  ||   | ||||| | | || ||   | | | ||||  | | | | | | ||
Db     167497 GCTGGCTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGTTGGTTG 167556

Qy        419 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 478
              ||  | | | | ||| | |||   ||| ||| | | |||||  || ||| ||| | | |
Db     167557 GCTGGTTGGTTGGCTGGTTGGTTGGCTGGCTGGTTGGCTGG-TTGGTTGGTTGGTTGGTT 167615

Qy        479 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 538
              |  || ||| ||| | |   |  || ||| |||   |   |  ||  || ||| | | |
Db     167616 GGGTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGCTGGCTGGTTGGCTGGTTGGTTGGTT 167675

Qy        539 GCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCT 598
              | ||| ||| |  |  | | |   || ||  || ||| |  |  | | |   |  || ||
Db     167676 GGCTGGTTGTTTGGTTGGCTGGCTGGTTGGCTGGTTGGTTGGTTGGCTGGTTGGTTGGCT 167735

Qy        599 GATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 658
              | | | ||| | |   || || | ||||| | |   |||| | || || | |   || |
Db     167736 GGTTGGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTT 167795

Qy        659 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCT 718
              |   || || | |   ||||| | || ||   |   || |    || || | | | || |
Db     167796 GGCTGGTTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGCTGGTTGGTTGGATGGTT 167855

Qy        719 GCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 778
              | | || || | |   |||  | ||||| | |   || ||  || ||   |   || |
Db     167856 GGTTGGTTGGTTGGTTGGCTGGTTGGCTGGTTGGTTGGTTGGCTGGTTGGCTGGTTGGTT 167915

Qy        779 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTT 838
              | | || || | |   |||  | ||||| |||   || ||  || | |  || ||| ||
Db     167916 GGTTGGATGGTTGGTTGGCTAGTTGGCTGGTAGGGTGGTTGGTTGGTTGACTGGTTGGTT 167975

Qy        839 GCTTGCTAG 847
              | ||| | |
Db     167976 GGTTGGTTG 167984


RESULT 95
AC103134
LOCUS       AC103134              238488 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-46K8, WORKING DRAFT SEQUENCE, 5
            unordered pieces.
ACCESSION   AC103134
VERSION     AC103134.8  GI:30579476
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 238488)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
```

```
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 238488)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 238488)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:23121862.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
               Center project name: GJHY
               Center clone name: CH230-46K8
            -------------- Summary Statistics
               Assembly program: Atlas 3.0;
               Consensus quality: 231556 bases at least Q40
               Consensus quality: 232783 bases at least Q30
               Consensus quality: 233921 bases at least Q20
               Estimated insert size: 244686; sum-of-contigs estimation
               Quality coverage: 9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   232888: contig of 232888 bp in length
            *   232889   232988: gap of unknown length
            *   232989   234664: contig of 1676 bp in length
            *   234665   234764: gap of unknown length
            *   234765   236085: contig of 1321 bp in length
            *   236086   236185: gap of unknown length
            *   236186   237259: contig of 1074 bp in length
            *   237260   237359: gap of unknown length
            *   237360   238488: contig of 1129 bp in length.
FEATURES            Location/Qualifiers
     source         1. .238488
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-46K8"
     misc_feature   1. .1820
                    /note="wgs_end_extension
                    clone_end:Sp6"
     misc_feature   1871. .3956
                    /note="wgs_end_extension
                    clone_end:Sp6"
     misc_feature   4857. .6437
                    /note="wgs_end_extension
                    clone_end:Sp6"
     misc_feature   complement(7762. .8503)
                    /note="clone_boundary
                    clone_end:Sp6
                    site:EcoRI
                    end_sequence:BH331187"
     misc_feature   229760. .232888
                    /note="wgs_contig"
     misc_feature   229822. .230347
                    /note="clone_boundary
                    clone_end:T7
                    site:EcoRI
                    end_sequence:BH331185"
BASE COUNT    63555 a  56184 c  54190 g  60847 t   3712 others
ORIGIN

  Query Match            14.9%;  Score 149.2;  DB 2;  Length 238488;
  Best Local Similarity   54.4%;  Pred. No. 4.7e-15;
  Matches  323;  Conservative    0;  Mismatches  268;  Indels    3;  Gaps    1;

Qy        282 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 341
              || | |  | || ||  || |   || | |  | || |   || |   || | | | |||
Db     132571 TGGTGGTGGTGGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTTGGCT 132630

Qy        342 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 401
              ||   ||| ||||    |   |  ||   | ||| | ||| |  ||   | | ||||||
Db     132631 GGTTGGCTGGCTGGCTAGCTGGCTGGCTGGTTGGTTGGCTGGCTGGTTGGTTGGCTGGCT 132690

Qy        402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461
                | | | | ||| |||||   | | ||| ||| |||||   | | | | ||| |||||
```

```
Db     132691 GGTTGGTTGGCTGGCTGGTTGGTTGGCTGGCTGGCTGGTTGGTTGGTTGGCTGGCTGGTT 132750

Qy        462 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 521
              ||| ||| | | |||||| |  ||| | | | | || ||| ||| | | | | |
Db     132751 GGCTAGCTGGTTAGCTGGCTGATTGTGTTGGTTGGTTGATTGGTTGGTTGGTTGGCTGGC 132810

Qy        522 TGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAG---GGCTGCTTGCTTGCTA 578
              || ||| ||| | | | | ||||| |||| || || | ||||| ||| ||| |
Db     132811 TGGTTGGTTGTTTGGTTGGCTGCTGGCTGGTTGGTTGGTTGGCTGGCTGGTTGGTTGGTT 132870

Qy        579 AGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 638
              |  | | |   |  ||  |   | ||||| |   || || | ||||| | |  ||
Db     132871 GGCTGGCTGGTTGGTTGGTTAGCTGGTTGATTGGTTGGTTGGTTGGCTGGTTGGTGGTTG 132930

Qy        639 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 698
              | | |||||   |   |||||  || ||   |   || || | |||||   |   || |
Db     132931 GGTTGGCTGGCCGGCTGGCTGGCTGGTTGGCTGGTTGGTTGGTTGGCTGACTGGCTGGTT 132990

Qy        699 CCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCT 758
              | || ||   |   ||||| | || ||   |   || |   |||||   |   || |
Db     132991 GGTTGGTTGGCTGGTTGGCTGGTTGGTTGGCTGGCTGGTTAGCTGGCTGGCTGGTTGGTT 133050

Qy        759 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 818
              | | || || | |   |||||  |||||  | |   || ||  |||||  |    || |
Db     133051 GGTTGGTTGGTTGGTTGGCTGGCTGGCTGGTTGGTTGGTTGGCTGGCTGGTTAGTTGGTT 133110

Qy        819 GCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGG 872
              | |||   | ||| ||   |  ||| |  ||  || | |    | | |||  ||
Db     133111 GGCTGGCTGGTTGGTTTTGTTGTTGTTGCTGCTGCTGTTGTTTTACCACTCTGG 133164
```

RESULT 96
AC119470/c

```
LOCUS       AC119470               231764 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-361I3, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC119470
VERSION     AC119470.5  GI:25099035
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 231764)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 231764)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 231764)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:23195740.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GVDR
                Center clone name: CH230-361I3
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 221446 bases at least Q40
                Consensus quality: 223410 bases at least Q30
                Consensus quality: 224649 bases at least Q20
                Estimated insert size: 227031; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
```

```
            * be preserved.
            *       1    19303: contig of 19303 bp in length
            *   19304    19403: gap of unknown length
            *   19404   228061: contig of 208658 bp in length
            *  228062   228161: gap of unknown length
            *  228162   229325: contig of 1164 bp in length
            *  229326   229425: gap of unknown length
            *  229426   231764: contig of 2339 bp in length.
FEATURES             Location/Qualifiers
     source          1. .231764
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-361I3"
     misc_feature    1. .992
                     /note="wgs_contig"
BASE COUNT    55064 a  59804 c  56805 g  53783 t   6308 others
ORIGIN

  Query Match             14.8%;  Score 148.4;  DB 2;  Length 231764;
  Best Local Similarity   57.2%;  Pred. No. 6.4e-15;
  Matches  269;  Conservative    0;  Mismatches  201;  Indels    0;  Gaps    0;

Qy        135 CCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGTGT 194
              ||| || ||| | ||  || |   | ||    || ||| ||||   || ||||     ||
Db     199302 CCCTGGCAGACCTGGCTATGGCACAGACTTACGGTGAGTGGGGCGTAGAGGACTTTCAGT 199243

Qy        195 GCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGG 254
              || ||| ||   | |  |  |   |  | |||||||||   ||  |||||||| |
Db     199242 GCCTGCCTGCCTGCCTGCCTGCCTGCTTCCCTGCCTGCCTGCCTGCCTGCTTCCCTGCCT 199183

Qy        255 GACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTG 314
              |       |||||||| |   |  |||||| |||      | | |   |||||| |||
Db     199182 GCTTCCCTGCCTGCTTCCCTCCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 199123

Qy        315 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTG 374
                | | |    ||||| ||| ||  |||||| |||      | | |   | ||||
Db     199122 GCCTGCCTGCCTCCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 199063

Qy        375 GCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 434
              || |||  |     |||| ||  |||| | ||| |||  ||  |||| | ||| |||
Db     199062 GCCTGCCTGCCTCCCTGCCTGCCTGCCTCCCTGCCTGCCTGCCTGCCTCCCTGCCTGCCT 199003

Qy        435 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 494
               || |||||| ||| |||  ||  |||||| | | |||  ||  |||| | ||| |||
Db     199002 CCTTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCTGCCTGCCTGCCTCCCTGCCTGCCT 198943

Qy        495 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 554
               |  |||||| ||| |||  ||  |||||| | | |||  ||   ||||| |||
Db     198942 CCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCTGCCTGCCTCCCTGCCTGCCTGCCT 198883

Qy        555 GACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGC 604
              | | |   | |||| ||| |||      | | |   |||||||||   ||
Db     198882 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGC 198833


RESULT 97
AC119267
LOCUS       AC119267              191953 bp    DNA     linear   HTG 24-FEB-2003
DEFINITION  Mus musculus clone RP24-322H19, WORKING DRAFT SEQUENCE, 14
            unordered pieces.
ACCESSION   AC119267
VERSION     AC119267.3  GI:28475980
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 191953)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-322H19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191953)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 191953)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 24, 2003 this sequence version replaced gi:28269582.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L25500
                Center clone name: 322_H_19
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 188248 bases at least Q40
                Consensus quality: 189506 bases at least Q30
                Consensus quality: 190207 bases at least Q20
                Insert size: 188000; agarose-fp
                Insert size: 190653; sum-of-contigs
                Quality coverage: 9.5 in Q20 bases; agarose-fp
                Quality coverage: 9.4 in Q20 bases; sum-of-contigs
            --------------
```

```
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     6538: contig of 6538 bp in length
            *    6539     6638: gap of 100 bp
            *    6639     7323: contig of 685 bp in length
            *    7324     7423: gap of 100 bp
            *    7424     8063: contig of 640 bp in length
            *    8064     8163: gap of 100 bp
            *    8164     9276: contig of 1113 bp in length
            *    9277     9376: gap of 100 bp
            *    9377    10779: contig of 1403 bp in length
            *   10780    10879: gap of 100 bp
            *   10880    12442: contig of 1563 bp in length
            *   12443    12542: gap of 100 bp
            *   12543    13692: contig of 1150 bp in length
            *   13693    13792: gap of 100 bp
            *   13793    16322: contig of 2530 bp in length
            *   16323    16422: gap of 100 bp
            *   16423    19255: contig of 2833 bp in length
            *   19256    19355: gap of 100 bp
            *   19356    84569: contig of 65214 bp in length
            *   84570    84669: gap of 100 bp
            *   84670   112524: contig of 27855 bp in length
            *  112525   112624: gap of 100 bp
            *  112625   134113: contig of 21489 bp in length
            *  134114   134213: gap of 100 bp
            *  134214   178370: contig of 44157 bp in length
            *  178371   178470: gap of 100 bp
            *  178471   191953: contig of 13483 bp in length.
FEATURES             Location/Qualifiers
     source          1. .191953
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-322H19"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .6538
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    6639. .7323
                     /note="assembly_fragment"
     misc_feature    7424. .8063
                     /note="assembly_fragment"
     misc_feature    8164. .9276
                     /note="assembly_fragment"
     misc_feature    9377. .10779
                     /note="assembly_fragment"
     misc_feature    10880. .12442
                     /note="assembly_fragment"
     misc_feature    12543. .13692
                     /note="assembly_fragment"
     misc_feature    13793. .16322
                     /note="assembly_fragment"
     misc_feature    16423. .19255
                     /note="assembly_fragment"
     misc_feature    19356. .84569
                     /note="assembly_fragment"
     misc_feature    84670. .112524
                     /note="assembly_fragment"
     misc_feature    112625. .134113
                     /note="assembly_fragment"
     misc_feature    134214. .178370
                     /note="assembly_fragment"
     misc_feature    178471. .191953
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    54331 a  41691 c  40641 g  53989 t   1301 others
ORIGIN

  Query Match             14.7%;  Score 147.4;  DB 2;  Length 191953;
  Best Local Similarity   69.8%;  Pred. No. 9.6e-15;
  Matches  199;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        269 TTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGC 328
              |||   ||| || ||||||||     |  ||   ||  ||| |||     |  |   ||
Db     146621 TTGTGTGCTTGCTTGCTTGCTTGCTTGCTAGCTAGCTAGCTAGCTAGCTAGCTTGCTTGC 146680

Qy        329 CTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGG 388
              |||||||| ||| || ||||||||     |   |   |  |||| |||||||| |   |
Db     146681 TTGCTTGCTAGCTAGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 146740

Qy        389 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 448
              |||| ||| ||  ||| ||| ||| ||| || ||||||||||||| ||| || |||||||
Db     146741 TTGCTTGCTTGCTAGCTAGCTAGCTAGCTAGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 146800

Qy        449 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 508
              ||||| ||| || |||||||||||| ||| ||  ||| ||||||| ||| || |||||||
Db     146801 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTAGCTAGCTAGCTTGCTTGCTTGCTTGCTTGC 146860

Qy        509 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGG 553
              ||||| ||| || |||||||||||| ||| || |||||   | ||
Db     146861 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTTGGGGGG 146905


RESULT 98
AC087842
LOCUS       AC087842               165608 bp    DNA     linear   HTG 31-JAN-2001
DEFINITION  Rattus norvegicus clone RP31-431A14 strain Brown Norway, WORKING
            DRAFT SEQUENCE, 8 unordered pieces.
ACCESSION   AC087842
VERSION     AC087842.1  GI:12621394
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 165608)
  AUTHORS   Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S.,
            Guan,X., Gupta,J., Ho,S.-L., Idol,J.R., Karlins,E., Lee-Lin,S.-Q.,
            Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B., Masiello,C.,
            Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,
            Shevchenko,Y., Snyder,B., Stantripop,S., Thomas,J.W., Thomas,P.J.,
            Tiongson,E.E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165608)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_mouse@nhgri.nih.gov
            -------------- Project Information
                 Center project name: ahd
                 Center clone name: 431A14
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 162851 bases at least Q40
                 Consensus quality: 163417 bases at least Q30
                 Consensus quality: 163699 bases at least Q20
```

```
                Insert size: 179000; agarose-fp
                Insert size: 121000; pulse-field-gel
                Insert size: 164908; sum-of-contigs
                Quality coverage: 9.61x in Q20 bases; agarose-fp
                Quality coverage: 14.22x in Q20 bases; pulse-field-gel
                Quality coverage: 10.43x in Q20 bases; sum-of-contigs
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4100: contig of 4100 bp in length
            *     4101     4200: gap of unknown length
            *     4201     8636: contig of 4436 bp in length
            *     8637     8736: gap of unknown length
            *     8737    14897: contig of 6161 bp in length
            *    14898    14997: gap of unknown length
            *    14998    22622: contig of 7625 bp in length
            *    22623    22722: gap of unknown length
            *    22723    39184: contig of 16462 bp in length
            *    39185    39284: gap of unknown length
            *    39285    64329: contig of 25045 bp in length
            *    64330    64429: gap of unknown length
            *    64430   105511: contig of 41082 bp in length
            *   105512   105611: gap of unknown length
            *   105612   165608: contig of 59997 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165608
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="Brown Norway"
                     /db_xref="taxon:10116"
                     /clone="RP31-431A14"
                     /clone_lib="RP31"
     misc_feature    1. .4100
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    4201. .8636
                     /note="assembly_fragment"
     misc_feature    8737. .14897
                     /note="assembly_fragment"
     misc_feature    14998. .22622
                     /note="assembly_fragment"
     misc_feature    22723. .39184
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    39285. .64329
                     /note="assembly_fragment"
     misc_feature    64430. .105511
                     /note="assembly_fragment"
     misc_feature    105612. .165608
                     /note="assembly_fragment"
BASE COUNT    42010 a  44676 c  42025 g  36150 t    747 others
ORIGIN

  Query Match             14.7%;  Score 146.8;  DB 2;  Length 165608;
  Best Local Similarity   56.4%;  Pred. No. 1.2e-14;
  Matches  274;  Conservative    0;  Mismatches  212;  Indels    0;  Gaps    0;

Qy        135 CCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGTGT 194
              ||| || ||| | ||  || |   | ||    || ||| ||||   || ||||     ||
Db      52859 CCCTGGCAGACCTGGCTATGGCACAGACTTACGGTGAGTGGGGCGTAGAGGACTTTCAGT 52918

Qy        195 GCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGG 254
              .|| ||| ||   | |  |  |   |  | |||||||||  ||  ||||||||  |
Db      52919 GCCTGCCTGCCTGCCTGCCTGCTTCCCTGCCTGCCTGCCTGCCTGCTTCCCTGCCT 52978

Qy        255 GACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTG 314
              |       ||||||||  |   |  ||||||  |||       | | |   ||||||  |||
Db      52979 GCTTCCCTGCCTGCTTCCCTCCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 53038

Qy        315 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTG 374
                | | |    |||||  |||  ||  ||||||  |||       | | |   | ||||
Db      53039 GCCTGCCTGCCTCCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 53098

Qy        375 GCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 434
              ||  |||  |     ||||  ||  |||| | |||  |||  ||  |||| | |||  |||
Db      53099 GCCTGCCTGCCTCCCTGCCTGCCTGCCTCCCTGCCTGCCTGCCTGCCTCCCTGCCTGCCT 53158

Qy        435 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 494
               ||  ||||||  |||  |||  ||  ||||||  | |  |||  ||  ||||  | |||  |||
Db      53159 CCTTGCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCTGCCTGCCTGCCTCCCTGCCTGCCT 53218

Qy        495 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 554
               |  ||||||  |||  |||  ||  ||||||  | | | |   |   |||  | |||
Db      53219 CCCTGCCTGCCTGCCTGCCTGCCTGCCTCCCTCCCTCCCTCCCTCCCTGCCTGCCT 53278

Qy        555 GACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAG 614
              | | |   | ||||  |||  |||       | | |   |||||||||   ||  ||   |
Db      53279 GCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCT 53338

Qy        615 GGCTGC 620
              | ||||
Db      53339 GCCTGC 53344


RESULT 99
AC119053/c
LOCUS       AC119053              165118 bp    DNA     linear   HTG 27-SEP-2002
DEFINITION  Danio rerio clone CH211-277I7 strain Tue, WORKING DRAFT SEQUENCE, 9
            ordered pieces.
ACCESSION   AC119053
VERSION     AC119053.2  GI:23334702
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 165118)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Cariaga,K., Coleman,B., Dietrich,N.L., Granite,S., Guan,X.,
            Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R.,
            Karlins,E., Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L.,
            Maduro,V.B., Margulies,E.H., Masiello,C., Maskeri,B.,
            Mastrian,S.D., McCloskey,J.C., McDowell,J., Paguirigan,C.,
            Pearson,R., Portnoy,M.E., Prasad,A., Reddix-Dugue,N.,
            Schueler,M.G., Sison,C., Stantripop,S., Thomas,J.W., Thomas,P.J.,
            Touchman,J.W., Vogt,J.L., Walker,M., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165118)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 165118)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Sep 27, 2002 this sequence version replaced gi:20279419.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
```

```
            Center project name: aqv
            Center clone name: 277I07

          The sequence data in this record represents an 'enhanced'
          version of a Phase 2 submission. Specifically, the indicated
          order and orientation of each sequence contig has been
          established using one or more of the following: read-pair
          data from individual subclones, overlaps with neighboring
          clones, alignment with available reference sequence (e.g.,
          human), and/or confirmation by PCR testing. In addition,
          the sequence assembly is based on at least 8X average
          coverage in Q20 bases and has been reviewed to rule out
          gross misassemblies, the low-quality ends of sequence
          contigs have been trimmed away, and each base is associated
          with a Phrap-derived quality score.
          -------------- Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 162840 bases at least Q40
            Consensus quality: 163595 bases at least Q30
            Consensus quality: 164091 bases at least Q20
            Insert size: 163000; agarose-fp
            Insert size: 164318; sum-of-contigs
            Quality coverage: 8.02x in Q20 bases; agarose-fp
            Quality coverage: 7.95x in Q20 bases; sum-of-contigs
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 9 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *        1    20544: contig of 20544 bp in length
          *    20545    20644: gap of unknown length
          *    20645    34251: contig of 13607 bp in length
          *    34252    34351: gap of unknown length
          *    34352    51803: contig of 17452 bp in length
          *    51804    51903: gap of unknown length
          *    51904    89342: contig of 37439 bp in length
          *    89343    89442: gap of unknown length
          *    89443    91814: contig of 2372 bp in length
          *    91815    91914: gap of unknown length
          *    91915    95203: contig of 3289 bp in length
          *    95204    95303: gap of unknown length
          *    95304   129870: contig of 34567 bp in length
          *   129871   129970: gap of unknown length
          *   129971   148921: contig of 18951 bp in length
          *   148922   149021: gap of unknown length
          *   149022   165118: contig of 16097 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165118
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /strain="Tue"
                     /db_xref="taxon:7955"
                     /clone="CH211-277I7"
                     /clone_lib="CH211"
     misc_feature    1. .20544
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    20645. .34251
                     /note="assembly_fragment"
     misc_feature    34352. .51803
                     /note="assembly_fragment"
     misc_feature    51904. .89342
                     /note="assembly_fragment"
     misc_feature    89443. .91814
                     /note="assembly_fragment"
     misc_feature    91915. .95203
                     /note="assembly_fragment"
     misc_feature    95304. .129870
                     /note="assembly_fragment"
     misc_feature    129971. .148921
                     /note="assembly_fragment"
     misc_feature    149022. .165118
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    51295 a  32475 c  30431 g  50116 t    801 others
ORIGIN

  Query Match             14.6%;  Score 146.4;  DB 2;  Length 165118;
  Best Local Similarity   52.1%;  Pred. No. 1.4e-14;
  Matches  353;  Conservative    0;  Mismatches  321;  Indels    4;  Gaps    1;

Qy        219 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 278
              || ||  ||  || | | | |  | |||| |  |  |   |   |  || ||| |  ||
Db      95201 GGTTGGTTGGTTGGTTGGTTGATTTCTTGGTTGGTTGGTTGGTTGGTTGGTTGATTTCTT 95142

Qy        279 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 338
              |  || ||| |  |  |    |   |  || ||| |  |  |   |   |  || ||| |
Db      95141 GGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 95082

Qy        339 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 398
              | | |  || ||| |     |   |   || || | | ||| | |   || || | | |
Db      95081 GGTTGGTTGGTTGATTTCTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 95022

Qy        399 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 458
              |  || ||| ||| | | | | ||| ||| ||| | | | |  || ||| || || | |
Db      95021 GGTTGGTTGGTTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTTGGTTGATTTCTTGGTT 94962

Qy        459 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 518
              |  || ||| ||| | | | |  || ||| ||| |  || |  || |||  || | | |
Db      94961 GGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGATTTCTTGGTTGGTTGGCTGGTTGGTT 94902

Qy        519 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTA 578
              |  | |||| ||| | | | |  | |||| |  |  | | |   || || ||| ||| |
Db      94901 GATTTCTTGGTTGGTTGGTTGATTTCTTGGTTGGTTGGCTGGTTGGTTGGTTGGTTGGTT 94842

Qy        579 AGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 638
               |  |   |    | ||  || | | |||||     || |||| || || |     || |
Db      94841 GGTTGGTTGATTTCTTGGTTGGTTGGTTGATTTGTTGGTTGCTTGGTTGATTTCTTGGTT 94782

Qy        639 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 698
              | | || || |      || || | || |||| |   |  |||| || |||| |   || |
Db      94781 GGTTGGTTGATTTGTTGGTTGGTTGGTTGCTTGCTTGCTTGCTTGGTTGCTTGGTTGGTT 94722

Qy        699 C----CTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 754
                  || || || | |   |||  | || || |     || |  |  | || | |
Db      94721 GGTTGCTTGGTTGGTTGGTTGGCTAGTTGGTTGGTTTGTTGGTTGGTTAGTTGGTTGGTC 94662

Qy        755 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 814
               | || | || |  |     || || |||| || |     || || | || || | |
Db      94661 TGTTGGTTGGTTAATTTCTTGGTTGGTGGGTTGATTACTTGGTTGGTTGGTTGGTTGGTT 94602

Qy        815 GGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCT 874
               | | || | | ||  || |||| ||| | |   |    |||| || | |   ||| |
Db      94601 AATTTCTTGGTTGGTTAATTTCTTGGTTGGTGGGCTGATTTCTTGGTTGGTTGGTGGGTT 94542

Qy        875 GCTGTGCAGCTGGGAGAA 892
              | | ||  | | ||  ||
Db      94541 GGTCTGTTGGTTGGTTAA 94524


RESULT 100
AC128150
LOCUS       AC128150              174032 bp    DNA     linear   HTG 20-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-386E19, *** SEQUENCING IN PROGRESS
            ***.
ACCESSION   AC128150
VERSION     AC128150.3  GI:25139403
```

```
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 174032)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174032)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 174032)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 20, 2002 this sequence version replaced gi:23803773.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GZOV
                Center clone name: CH230-386E19
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 159301 bases at least Q40
                Consensus quality: 160205 bases at least Q30
                Consensus quality: 160799 bases at least Q20
                Estimated insert size: 165067; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   174032: contig of 174032 bp in length.
FEATURES             Location/Qualifiers
     source          1. .174032
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-386E19"
     misc_feature    complement(163561. .164704)
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:RXBAE34TJB"
     misc_feature    complement(163589. .164704)
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:RXBAE34TJC"
     misc_feature    165382. .167144
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    167195. .168910
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    172327. .174032
                     /note="wgs_end_extension
                     clone_end:T7"
BASE COUNT    44840 a  32463 c  33151 g  50641 t  12937 others
ORIGIN

  Query Match            14.6%;  Score 146.4;  DB 2;  Length 174032;
  Best Local Similarity   58.5%;  Pred. No. 1.4e-14;
  Matches  275;  Conservative    0;  Mismatches  191;  Indels    4;  Gaps    1;

Qy        222 TGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 281
              ||  |||  | | |||||  |||| | |    |||  |  ||  |||| | | |||||
Db      25636 TGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTAGGTTGCTGGAT 25695
```

```
Qy        282 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 341
              |||| | | |  |||  |  ||  |||| | |    | |  |  |   ||||   | |||
Db      25696 TGCTGGATTACTGGATTGCTGGATTGCTGGATTGCTGCATTGCTGCATTGCTGCATTGCT 25755

Qy        342 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 401
              ||  |||| | |    |||  |  || ||||||| ||||| |       | | ||||| |
Db      25756 GGATTGCTGGATTGCTGGATTGCTGGACTGCTGGATTGCTGGATTACTGGATTGCTGGAC 25815

Qy        402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461
              |||| | ||||| | | ||  | |||||  |   |||||  ||||  |  ||||  | | ||
Db      25816 TGCTGGATTGCTGGATTGCTGGATTGCTGGACTGCTGGATTGCTGGACTGCTGGATTGCT 25875

Qy        462 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 521
              | || || | | |||||  |||| |  |||| | | ||  | || || | | |||||  |
Db      25876 GGATTACTGGATTGCTGGATTGCTGGACTGCTGGATTGCTGGATTACTGGATTGCTGGAC 25935

Qy        522 TGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGG 581
              |||| | ||||| | | ||  | ||||||     |   |   |   |  |||| | |
Db      25936 TGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGACTGCTGGATTGCT 25995

Qy        582 GGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCT 641
              ||||  |   |  |||  ||  ||| ||| |  || |||||||  ||||    ||  ||||
Db      25996 GGACTGCTGGATTGCTGGACTGCTGGATTGCTGGACTGCTGGATTGCT----GGATTGCT 26051

Qy        642 GGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAG 691
              ||  |||| |   |   | | |   | | |  ||  |||||| |||||||
Db      26052 GGATTGCTGGACTGCTGGATTGCTGGATTGCTGGATTGCTGGACTGCTAG 26101

Search completed: August 21, 2003, 23:30:22
Job time : 3752.36 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 21, 2003, 21:03:57 ; Search time 76.038 Seconds
(without alignments)
5810.577 Million cell updates/sec

Title: US-09-831-000-1_COPY_132719_133719
Perfect score: 1001
Sequence: 1 cgccgcccggcatggcggtc.....cgggcacccggcccagatct 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	127.6	12.7	35100	2	US-08-770-379-17	Sequence 17, Appl
c	2	127.6	12.7	35100	3	US-08-757-669A-17	Sequence 17, Appl
c	3	127.6	12.7	35100	4	US-09-230-371A-17	Sequence 17, Appl
c	4	109.8	11.0	11558	5	PCT-US93-06251-23	Sequence 23, Appl
	5	77.4	7.7	1830121	4	US-09-557-884-1	Sequence 1, Appli
	6	77.4	7.7	1830121	4	US-09-643-990A-1	Sequence 1, Appli
c	7	64.2	6.4	1771	2	US-08-533-669A-7	Sequence 7, Appli
c	8	64.2	6.4	1771	2	US-08-511-872-1	Sequence 1, Appli
c	9	64.2	6.4	1771	4	US-09-183-861-7	Sequence 7, Appli
c	10	64.2	6.4	1771	4	US-09-022-765-7	Sequence 7, Appli
c	11	64.2	6.4	1771	4	US-09-551-974A-7	Sequence 7, Appli
c	12	58	5.8	3252	4	US-08-476-102A-4	Sequence 4, Appli
c	13	58	5.8	3252	5	PCT-US96-05320A-714	Sequence 714, App
c	14	56	5.6	45546	4	US-09-146-053-6	Sequence 6, Appli
c	15	55.6	5.6	1960	2	US-08-595-937A-1	Sequence 1, Appli
c	16	52.6	5.3	319608	4	US-09-539-333D-1	Sequence 1, Appli
c	17	52.6	5.3	319608	4	US-09-679-409-1	Sequence 1, Appli
c	18	52.4	5.2	350	2	US-08-332-766A-20	Sequence 20, Appl
	19	51.8	5.2	4403	2	US-08-284-941-1	Sequence 1, Appli
	20	51.8	5.2	4403	2	US-08-447-642-1	Sequence 1, Appli
	21	51.8	5.2	4403	3	US-09-236-503-1	Sequence 1, Appli
	22	51.8	5.2	4403	5	PCT-US93-02147A-1	Sequence 1, Appli
c	23	51.6	5.2	538	2	US-08-332-766A-24	Sequence 24, Appl
c	24	51.4	5.1	72604	4	US-09-268-992-7	Sequence 7, Appli
c	25	51.4	5.1	72604	4	US-09-657-474-7	Sequence 7, Appli
	26	50.4	5.0	405	2	US-08-903-800A-3	Sequence 3, Appli
c	27	49.8	5.0	348	2	US-08-623-906A-14	Sequence 14, Appl
c	28	49.2	4.9	289	3	US-09-007-005-17	Sequence 17, Appl
c	29	49.2	4.9	289	3	US-09-244-796-17	Sequence 17, Appl
c	30	48.8	4.9	650	1	US-08-229-515A-13	Sequence 13, Appl
c	31	48.8	4.9	650	1	US-08-645-865-13	Sequence 13, Appl
c	32	48.8	4.9	1682	3	US-09-318-443-7	Sequence 7, Appli
	33	47.8	4.8	3228	4	US-09-579-664B-5	Sequence 5, Appli
	34	47.8	4.8	7218	1	US-08-232-463-14	Sequence 14, Appl
c	35	47.6	4.8	852	4	US-09-252-991A-10276	Sequence 10276, A
c	36	47.6	4.8	2778	4	US-09-252-991A-10486	Sequence 10486, A
	37	47.6	4.8	3339	4	US-09-252-991A-10685	Sequence 10685, A
	38	47.4	4.7	72604	4	US-09-268-992-7	Sequence 7, Appli
	39	47.4	4.7	72604	4	US-09-657-474-7	Sequence 7, Appli
	40	47.2	4.7	320	3	US-09-165-264-7	Sequence 7, Appli
	41	47.2	4.7	320	3	US-09-165-264-13	Sequence 13, Appl
c	42	46.6	4.7	1120	3	US-08-586-165-1	Sequence 1, Appli
	43	46.6	4.7	1393	2	US-08-765-875-1	Sequence 1, Appli
	44	46.6	4.7	1393	3	US-08-795-671-1	Sequence 1, Appli
	45	46.6	4.7	1393	4	US-09-454-540-1	Sequence 1, Appli
c	46	46.6	4.7	1978	1	US-07-753-520B-2	Sequence 2, Appli
c	47	46.6	4.7	3937	3	US-08-586-165-8	Sequence 8, Appli
c	48	46.6	4.7	9115	1	US-07-753-520B-3	Sequence 3, Appli
c	49	46.6	4.7	152331	3	US-09-128-155-16	Sequence 16, Appl
c	50	46.4	4.6	124	2	US-08-332-766A-37	Sequence 37, Appl
	51	46.4	4.6	3507	2	US-08-775-009-36	Sequence 36, Appl
	52	46.2	4.6	320	3	US-09-165-264-11	Sequence 11, Appl
c	53	46.2	4.6	951	4	US-09-252-991A-16386	Sequence 16386, A
	54	46.2	4.6	1092	4	US-09-252-991A-16212	Sequence 16212, A
c	55	46.2	4.6	1311	4	US-09-252-991A-16493	Sequence 16493, A
	56	46.2	4.6	1530	4	US-09-252-991A-16005	Sequence 16005, A
c	57	46.2	4.6	1877	4	US-09-780-173A-10	Sequence 10, Appl
	58	46.2	4.6	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	59	46.2	4.6	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	60	45.8	4.6	318	3	US-09-165-264-12	Sequence 12, Appl
	61	45.8	4.6	90541	4	US-09-759-359A-3	Sequence 3, Appli
	62	45.6	4.6	319	3	US-09-165-264-8	Sequence 8, Appli
c	63	45.6	4.6	1559	3	US-09-019-095A-7	Sequence 7, Appli
c	64	45.4	4.5	23187	4	US-09-499-522-1	Sequence 1, Appli
c	65	45	4.5	275	4	US-08-585-593A-42	Sequence 42, Appl
	66	45	4.5	320	3	US-09-165-264-14	Sequence 14, Appl
	67	45	4.5	6453	1	US-08-306-691B-14	Sequence 14, Appl
	68	45	4.5	6453	3	US-09-209-668-10	Sequence 10, Appl
	69	45	4.5	6453	3	US-09-356-952-8	Sequence 8, Appli
c	70	44.8	4.5	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	71	44.8	4.5	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	72	44.2	4.4	227	4	US-09-056-556-165	Sequence 165, App
	73	44.2	4.4	227	4	US-09-072-596-160	Sequence 160, App
	74	44.2	4.4	400	4	US-09-056-556-179	Sequence 179, App
	75	44.2	4.4	400	4	US-09-072-596-174	Sequence 174, App
	76	44.2	4.4	176373	3	US-09-128-155-17	Sequence 17, Appl
	77	44	4.4	801	2	US-08-770-379-16	Sequence 16, Appl
	78	44	4.4	801	3	US-08-757-669A-16	Sequence 16, Appl
	79	44	4.4	801	4	US-09-230-371A-16	Sequence 16, Appl
c	80	44	4.4	12001	1	US-08-458-568A-11	Sequence 11, Appl
	81	43.8	4.4	1413	4	US-08-984-709A-52	Sequence 52, Appl
c	82	43.8	4.4	1667	1	US-08-485-284A-1	Sequence 1, Appli
	83	43.6	4.4	756	4	US-09-252-991A-6106	Sequence 6106, Ap
	84	43.6	4.4	912	4	US-09-252-991A-6186	Sequence 6186, Ap
c	85	43.6	4.4	915	4	US-09-252-991A-6445	Sequence 6445, Ap
c	86	43.6	4.4	1200	4	US-09-252-991A-6282	Sequence 6282, Ap
	87	43.6	4.4	1806	4	US-09-252-991A-6023	Sequence 6023, Ap
c	88	43.6	4.4	20966	4	US-09-776-976-7	Sequence 7, Appli
c	89	43.6	4.4	20966	4	US-09-909-547-7	Sequence 7, Appli
c	90	43.6	4.4	20966	4	US-09-569-852B-1	Sequence 1, Appli
c	91	43.2	4.3	892	3	US-09-179-558-64	Sequence 64, Appl
c	92	43.2	4.3	892	4	US-09-722-825-64	Sequence 64, Appl
c	93	43.2	4.3	892	4	US-09-722-487-64	Sequence 64, Appl
c	94	43.2	4.3	892	4	US-09-722-708-64	Sequence 64, Appl
c	95	43.2	4.3	936	3	US-09-179-558-62	Sequence 62, Appl
c	96	43.2	4.3	936	4	US-09-722-825-62	Sequence 62, Appl
c	97	43.2	4.3	936	4	US-09-722-487-62	Sequence 62, Appl
c	98	43.2	4.3	936	4	US-09-722-708-62	Sequence 62, Appl
	99	43.2	4.3	2888	4	US-08-765-907A-1	Sequence 1, Appli
	100	43.2	4.3	4257	2	US-08-690-473-1	Sequence 1, Appli

	101	43.2	4.3	4257	3	US-09-259-821A-1	Sequence 1, Appli
	102	43.2	4.3	4257	3	US-08-843-659-1	Sequence 1, Appli
	103	43	4.3	1641	4	US-09-252-991A-2066	Sequence 2066, Ap
c	104	43	4.3	2112	4	US-09-252-991A-1820	Sequence 1820, Ap
	105	43	4.3	2283	4	US-09-252-991A-2140	Sequence 2140, Ap
c	106	43	4.3	18443	3	US-09-078-294-6	Sequence 6, Appli
c	107	42.8	4.3	1461	1	US-08-450-360-1	Sequence 1, Appli
	108	42.8	4.3	4223	3	US-08-845-258-7	Sequence 7, Appli
c	109	42.8	4.3	4223	3	US-08-845-258-45	Sequence 45, Appl
	110	42.8	4.3	4223	3	US-08-990-571-7	Sequence 7, Appli
c	111	42.8	4.3	4223	3	US-08-990-571-45	Sequence 45, Appl
	112	42.8	4.3	4223	4	US-08-723-142A-7	Sequence 7, Appli
c	113	42.8	4.3	4223	4	US-08-723-142A-45	Sequence 45, Appl
	114	42.8	4.3	4223	4	US-09-528-784A-7	Sequence 7, Appli
c	115	42.8	4.3	4223	4	US-09-528-784A-45	Sequence 45, Appl
	116	42.8	4.3	4223	4	US-09-569-098A-7	Sequence 7, Appli
c	117	42.8	4.3	4223	4	US-09-569-098A-45	Sequence 45, Appl
c	118	42.6	4.3	2124	4	US-09-266-965-44	Sequence 44, Appl
c	119	42.6	4.3	2991	3	US-08-795-430-48	Sequence 48, Appl
c	120	42.6	4.3	2991	4	US-09-355-700-48	Sequence 48, Appl
	121	42.6	4.3	53500	4	US-09-266-965-76	Sequence 76, Appl
c	122	42.6	4.3	71989	4	US-09-443-501A-2	Sequence 2, Appli
c	123	42.4	4.2	4257	2	US-08-690-473-1	Sequence 1, Appli
c	124	42.4	4.2	4257	3	US-09-259-821A-1	Sequence 1, Appli
c	125	42.4	4.2	4257	3	US-08-843-659-1	Sequence 1, Appli
c	126	42.4	4.2	4695	2	US-08-231-193A-57	Sequence 57, Appl
c	127	42.4	4.2	4695	2	US-08-486-273A-57	Sequence 57, Appl
c	128	42.4	4.2	4695	3	US-08-940-086A-57	Sequence 57, Appl
c	129	42.4	4.2	4695	4	US-08-940-035A-57	Sequence 57, Appl
c	130	42.4	4.2	4695	4	US-08-935-105A-57	Sequence 57, Appl
c	131	42.4	4.2	4695	4	US-09-648-797-57	Sequence 57, Appl
c	132	42.4	4.2	4695	4	US-09-386-123-57	Sequence 57, Appl
	133	42.4	4.2	12001	1	US-08-458-568A-11	Sequence 11, Appl
	134	42.4	4.2	162450	4	US-09-345-882-1	Sequence 1, Appli
	135	42.2	4.2	4405	1	US-07-885-972A-3	Sequence 3, Appli
	136	42.2	4.2	4405	2	US-08-745-880-3	Sequence 3, Appli
	137	42.2	4.2	4405	2	US-08-480-382-3	Sequence 3, Appli
	138	42	4.2	460	2	US-08-903-800A-4	Sequence 4, Appli
	139	42	4.2	1561	4	US-08-765-907A-11	Sequence 11, Appl
c	140	42	4.2	2445	3	US-09-298-367B-9	Sequence 9, Appli
c	141	42	4.2	3000	3	US-08-460-269C-5	Sequence 5, Appli
c	142	42	4.2	8501	3	US-09-298-367B-6	Sequence 6, Appli
c	143	41.8	4.2	1142	4	US-08-478-486F-11	Sequence 11, Appl
c	144	41.8	4.2	1219	5	PCT-US93-06251-11	Sequence 11, Appl
	145	41.8	4.2	3855	3	US-08-974-549A-4	Sequence 4, Appli
	146	41.8	4.2	3855	4	US-08-912-951-4	Sequence 4, Appli
	147	41.8	4.2	4015	3	US-08-851-843A-224	Sequence 224, App
	148	41.8	4.2	4015	3	US-08-974-549A-1	Sequence 1, Appli
	149	41.8	4.2	4015	3	US-08-854-050-224	Sequence 224, App
	150	41.8	4.2	4015	4	US-09-430-323-224	Sequence 224, App
	151	41.8	4.2	4015	4	US-09-572-423B-3	Sequence 3, Appli
	152	41.8	4.2	4015	4	US-09-128-354-1	Sequence 1, Appli
	153	41.8	4.2	4015	4	US-09-675-321-1	Sequence 1, Appli
	154	41.8	4.2	4015	4	US-09-052-919-1	Sequence 1, Appli
	155	41.8	4.2	4015	4	US-08-912-951-1	Sequence 1, Appli
	156	41.8	4.2	4015	4	US-09-733-294A-3	Sequence 3, Appli
	157	41.8	4.2	4037	3	US-08-974-549A-343	Sequence 343, App
	158	41.8	4.2	4200	4	US-08-912-951-6	Sequence 6, Appli
	159	41.8	4.2	4335	3	US-08-974-549A-6	Sequence 6, Appli
	160	41.8	4.2	15418	4	US-09-783-203-1	Sequence 1, Appli
	161	41.8	4.2	51552	4	US-09-733-294A-30	Sequence 30, Appl
	162	41.6	4.2	1845	4	US-09-614-034-188	Sequence 188, App
	163	41.6	4.2	2126	3	US-09-237-543-1	Sequence 1, Appli
	164	41.6	4.2	2126	4	US-09-644-450-1	Sequence 1, Appli
	165	41.4	4.1	366	4	US-09-252-991A-10769	Sequence 10769, A
c	166	41.4	4.1	446	2	US-08-332-766A-26	Sequence 26, Appl
	167	41.4	4.1	1438	4	US-08-928-069-5	Sequence 5, Appli
	168	41.4	4.1	1438	4	US-08-828-683A-5	Sequence 5, Appli
	169	41.4	4.1	64467	4	US-09-803-671B-3	Sequence 3, Appli
c	170	41.2	4.1	471	4	US-09-252-991A-263	Sequence 263, App
c	171	41.2	4.1	813	4	US-09-252-991A-273	Sequence 273, App
c	172	41.2	4.1	837	4	US-09-252-991A-288	Sequence 288, App
	173	41.2	4.1	1017	4	US-09-252-991A-249	Sequence 249, App
	174	41.2	4.1	1524	4	US-09-252-991A-259	Sequence 259, App
c	175	41.2	4.1	8147	4	US-09-514-247A-9	Sequence 9, Appli
	176	41	4.1	1227	2	US-08-903-800A-1	Sequence 1, Appli
	177	41	4.1	47981	4	US-09-679-279-1	Sequence 1, Appli
c	178	40.8	4.1	618	4	US-09-252-991A-10230	Sequence 10230, A
c	179	40.8	4.1	1089	4	US-09-252-991A-10346	Sequence 10346, A
	180	40.8	4.1	1395	4	US-09-252-991A-10801	Sequence 10801, A
c	181	40.8	4.1	1506	4	US-09-252-991A-1159	Sequence 1159, Ap
	182	40.8	4.1	2667	4	US-09-252-991A-1396	Sequence 1396, Ap
c	183	40.8	4.1	2985	4	US-09-252-991A-1210	Sequence 1210, Ap
	184	40.8	4.1	3978	3	US-08-726-214-1	Sequence 1, Appli
	185	40.8	4.1	4029	3	US-08-851-843A-173	Sequence 173, App
	186	40.8	4.1	4029	3	US-08-974-549A-292	Sequence 292, App
	187	40.8	4.1	4029	3	US-08-854-050-173	Sequence 173, App
	188	40.8	4.1	4029	4	US-09-430-323-173	Sequence 173, App
	189	40.8	4.1	5707	2	US-08-472-809B-8	Sequence 8, Appli
	190	40.8	4.1	6345	2	US-08-472-809B-7	Sequence 7, Appli
	191	40.8	4.1	7898	4	US-08-984-709A-49	Sequence 49, Appl
	192	40.8	4.1	19650	4	US-09-819-989-3	Sequence 3, Appli
c	193	40.6	4.1	248	3	US-09-007-005-32	Sequence 32, Appl
c	194	40.6	4.1	248	3	US-09-244-796-32	Sequence 32, Appl
c	195	40.6	4.1	277	3	US-09-007-005-3	Sequence 3, Appli
c	196	40.6	4.1	277	3	US-09-244-796-3	Sequence 3, Appli
	197	40.6	4.1	1257	4	US-09-252-991A-15817	Sequence 15817, A
c	198	40.6	4.1	1506	4	US-09-252-991A-15902	Sequence 15902, A
c	199	40.6	4.1	2013	4	US-09-252-991A-15874	Sequence 15874, A
c	200	40.6	4.1	3177	1	US-08-042-747A-4	Sequence 4, Appli
	201	40.4	4.0	453	2	US-08-903-800A-2	Sequence 2, Appli
	202	40.4	4.0	1679	2	US-08-365-486A-11	Sequence 11, Appl
	203	40.4	4.0	1679	3	US-08-880-342-11	Sequence 11, Appl
c	204	40.4	4.0	4848	4	US-08-955-957A-1	Sequence 1, Appli
c	205	40.4	4.0	4848	4	US-08-955-957A-4	Sequence 4, Appli
c	206	40.4	4.0	4848	4	US-08-955-957A-6	Sequence 6, Appli
	207	40.2	4.0	1776	1	US-08-722-001-29	Sequence 29, Appl
	208	40.2	4.0	2140	1	US-08-334-698-1	Sequence 1, Appli
	209	40.2	4.0	2140	1	US-08-228-932-1	Sequence 1, Appli
	210	40.2	4.0	2140	1	US-08-468-939-1	Sequence 1, Appli
	211	40.2	4.0	2140	2	US-08-406-855A-1	Sequence 1, Appli
	212	40.2	4.0	2140	2	US-08-722-190-1	Sequence 1, Appli
	213	40.2	4.0	2140	3	US-08-244-354-1	Sequence 1, Appli
	214	40.2	4.0	2140	3	US-09-206-899-1	Sequence 1, Appli
	215	40.2	4.0	2140	4	US-09-444-783-1	Sequence 1, Appli
	216	40.2	4.0	2140	4	US-09-688-415-1	Sequence 1, Appli
	217	40.2	4.0	2140	5	PCT-US95-04203-1	Sequence 1, Appli
c	218	40.2	4.0	5253	4	US-09-423-890-7	Sequence 7, Appli
	219	40	4.0	402	3	US-09-209-915-1	Sequence 1, Appli
	220	40	4.0	402	4	US-09-209-916-2	Sequence 2, Appli
c	221	40	4.0	1756	4	US-09-016-434-1197	Sequence 1197, Ap
c	222	40	4.0	83450	4	US-09-811-469-3	Sequence 3, Appli
c	223	40	4.0	202001	4	US-09-734-674-3	Sequence 3, Appli
	224	39.8	4.0	530	3	US-08-758-662-4	Sequence 4, Appli
c	225	39.8	4.0	815	3	US-08-818-112-139	Sequence 139, App
c	226	39.8	4.0	815	4	US-08-818-111-134	Sequence 134, App
c	227	39.8	4.0	815	4	US-09-056-556-139	Sequence 139, App
c	228	39.8	4.0	815	4	US-09-072-596-134	Sequence 134, App
	229	39.8	4.0	4524	2	US-08-845-998-7	Sequence 7, Appli
	230	39.8	4.0	4524	3	US-09-206-537-7	Sequence 7, Appli
	231	39.8	4.0	4524	3	US-09-430-854-7	Sequence 7, Appli
c	232	39.6	4.0	465	4	US-09-252-991A-16275	Sequence 16275, A
c	233	39.6	4.0	1512	3	US-08-476-509B-1	Sequence 1, Appli
c	234	39.6	4.0	1626	3	US-08-348-518C-1	Sequence 1, Appli
c	235	39.6	4.0	2150	2	US-08-318-837-1	Sequence 1, Appli
	236	39.6	4.0	11219	1	US-07-642-734C-1	Sequence 1, Appli
	237	39.6	4.0	11219	3	US-08-439-009A-1	Sequence 1, Appli
c	238	39.4	3.9	1550	2	US-08-609-443B-17	Sequence 17, Appl
c	239	39.4	3.9	1550	4	US-08-851-896-17	Sequence 17, Appl
	240	39.4	3.9	2022	4	US-08-687-691B-3	Sequence 3, Appli
c	241	39.4	3.9	2575	4	US-09-620-312D-371	Sequence 371, App
c	242	39.4	3.9	4897	6	5196516-7	Patent No. 5196516
c	243	39.2	3.9	537	3	US-08-149-101A-6	Sequence 6, Appli
c	244	39.2	3.9	537	5	PCT-US94-12873-6	Sequence 6, Appli
c	245	39.2	3.9	795	3	US-08-713-556F-41	Sequence 41, Appl
c	246	39.2	3.9	801	2	US-08-770-379-16	Sequence 16, Appl

c 247	39.2	3.9	801	3	US-08-757-669A-16	Sequence 16, Appl
c 248	39.2	3.9	801	4	US-09-298-568-3	Sequence 3, Appli
c 249	39.2	3.9	801	4	US-09-230-371A-16	Sequence 16, Appl
c 250	39.2	3.9	879	4	US-09-252-991A-12575	Sequence 12575, A
c 251	39.2	3.9	885	4	US-09-252-991A-12721	Sequence 12721, A
c 252	39.2	3.9	977	1	US-08-017-522A-3	Sequence 3, Appli
c 253	39.2	3.9	977	1	US-08-115-680-1	Sequence 1, Appli
c 254	39.2	3.9	977	1	US-07-941-372-1	Sequence 1, Appli
c 255	39.2	3.9	977	5	PCT-US93-08247-1	Sequence 1, Appli
c 256	39.2	3.9	985	6	5215895-1	Patent No. 5215895
c 257	39.2	3.9	1208	1	US-07-949-516A-3	Sequence 3, Appli
c 258	39.2	3.9	1208	2	US-08-814-459-3	Sequence 3, Appli
c 259	39.2	3.9	1208	3	US-09-122-525-3	Sequence 3, Appli
c 260	39.2	3.9	1312	1	US-08-205-506A-1	Sequence 1, Appli
c 261	39.2	3.9	1312	5	PCT-US94-02389-1	Sequence 1, Appli
262	39.2	3.9	2145	4	US-09-252-991A-13235	Sequence 13235, A
c 263	39.2	3.9	3632	1	US-07-745-382-13	Sequence 13, Appl
c 264	39.2	3.9	3632	1	US-07-921-848-13	Sequence 13, Appl
c 265	39.2	3.9	3632	1	US-08-115-680-3	Sequence 3, Appli
c 266	39.2	3.9	3632	1	US-07-941-372-3	Sequence 3, Appli
c 267	39.2	3.9	3632	1	US-08-165-301A-13	Sequence 13, Appl
c 268	39.2	3.9	3632	3	US-08-810-436-13	Sequence 13, Appl
c 269	39.2	3.9	3632	5	PCT-US93-08247-3	Sequence 3, Appli
c 270	39.2	3.9	3632	5	PCT-US94-14179-13	Sequence 13, Appl
271	39.2	3.9	3646	4	US-09-300-958A-34	Sequence 34, Appl
c 272	39.2	3.9	4114	1	US-08-165-301A-29	Sequence 29, Appl
c 273	39.2	3.9	4114	3	US-08-810-436-29	Sequence 29, Appl
c 274	39.2	3.9	4114	5	PCT-US94-14179-29	Sequence 29, Appl
c 275	39.2	3.9	5993	3	US-09-383-630-1	Sequence 1, Appli
c 276	39.2	3.9	5993	3	US-09-383-630-2	Sequence 2, Appli
c 277	39.2	3.9	15144	3	US-08-458-434A-6	Sequence 6, Appli
278	39.2	3.9	20235	1	US-07-642-734C-3	Sequence 3, Appli
279	39.2	3.9	20235	3	US-08-439-009A-3	Sequence 3, Appli
280	39.2	3.9	71989	4	US-09-443-501A-2	Sequence 2, Appli
281	39	3.9	678	4	US-09-252-991A-13642	Sequence 13642, A
c 282	39	3.9	1247	1	US-08-278-729A-32	Sequence 32, Appl
c 283	39	3.9	1247	1	US-08-155-343A-32	Sequence 32, Appl
c 284	39	3.9	1247	1	US-08-406-672-32	Sequence 32, Appl
c 285	39	3.9	1247	1	US-08-643-563A-32	Sequence 32, Appl
c 286	39	3.9	1247	1	US-08-643-763A-32	Sequence 32, Appl
c 287	39	3.9	1247	1	US-08-462-623-32	Sequence 32, Appl
c 288	39	3.9	1247	1	US-08-451-953A-32	Sequence 32, Appl
c 289	39	3.9	1247	2	US-08-445-468A-32	Sequence 32, Appl
c 290	39	3.9	1247	2	US-08-461-397A-32	Sequence 32, Appl
c 291	39	3.9	1247	2	US-08-912-088-32	Sequence 32, Appl
c 292	39	3.9	1247	3	US-08-278-730A-32	Sequence 32, Appl
c 293	39	3.9	1247	3	US-08-445-467-32	Sequence 32, Appl
c 294	39	3.9	1247	3	US-08-480-515A-32	Sequence 32, Appl
c 295	39	3.9	1247	4	US-09-170-936-32	Sequence 32, Appl
c 296	39	3.9	1247	4	US-08-461-113-32	Sequence 32, Appl
c 297	39	3.9	1247	4	US-08-456-033-32	Sequence 32, Appl
c 298	39	3.9	1247	4	US-08-643-321-30	Sequence 30, Appl
c 299	39	3.9	1247	5	PCT-US93-07190-32	Sequence 32, Appl
c 300	39	3.9	1247	5	PCT-US93-07231-32	Sequence 32, Appl

ALIGNMENTS

```
RESULT 1
US-08-770-379-17/c
; Sequence 17, Application US/08770379
; Patent No. 5849564
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
;    TITLE OF INVENTION:  HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,379
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  52342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  35100 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-770-379-17

  Query Match             12.7%;  Score 127.6;  DB 2;  Length 35100;
  Best Local Similarity   66.9%;  Pred. No. 1.7e-17;
  Matches  198;  Conservative    0;  Mismatches   94;  Indels    4;  Gaps    1;

Qy        595 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 654
              | ||  |||  || |    ||||| ||||| ||||    ||||| ||||| ||||    |
Db      24611 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24552

Qy        655 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 714
              |||| ||||| ||||    ||||| ||||| ||||    |||||||||| ||||    |
Db      24551 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24492

Qy        715 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 774
              |||| ||||| ||||    |||||||||| ||||    ||||| ||||| ||||    |
Db      24491 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24432

Qy        775 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT----GCCTGCTGGCTT 830
              |||| ||||| ||||    ||||| ||||| ||||    |||||    |  ||||| |
Db      24431 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24372

Qy        831 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTG 886
              | | || |  ||||    ||||| || |  |||| |  ||||| ||| |  ||||
Db      24371 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTG 24316


RESULT 2
US-08-757-669A-17/c
; Sequence 17, Application US/08757669A
; Patent No. 6183751
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
;    TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
```

```
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/757,669A
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  45185-F
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  35100 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-757-669A-17

  Query Match             12.7%;  Score 127.6;  DB 3;  Length 35100;
  Best Local Similarity   66.9%;  Pred. No. 1.7e-17;
  Matches  198;  Conservative    0;  Mismatches   94;  Indels    4;  Gaps    1;

Qy        595 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 654
              | ||  |||  || |    ||||| ||||| ||||    ||||| ||||| ||||    |
Db      24611 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24552

Qy        655 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 714
              |||| ||||| ||||    ||||| ||||| ||||    |||||||||| ||||    |
Db      24551 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24492

Qy        715 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 774
              |||| ||||| ||||    |||||||||| ||||    ||||| ||||| ||||    |
Db      24491 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24432

Qy        775 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT----GCCTGCTGGCTT 830
              |||| ||||| ||||    ||||| ||||| ||||    |||||    |  ||||| |
Db      24431 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24372

Qy        831 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTG 886
              |  | || |  ||||   |||||  || |  |||| |  ||||| ||| |  ||||
Db      24371 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTG 24316


RESULT 3
US-09-230-371A-17/c
; Sequence 17, Application US/09230371A
; Patent No. 6348586
; GENERAL INFORMATION:
;  APPLICANT:  Chang, Yuan
;  APPLICANT:  Bohenzky, Roy A
;  APPLICANT:  Russo, James J
;  APPLICANT:  Edelman, Isidore S
;  APPLICANT:  Moore, Patrick S
;  TITLE OF INVENTION:  UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE:  45185-G-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/230,371A
;  CURRENT FILING DATE:  1999-11-17
;  PRIOR APPLICATION NUMBER:  PCT/US97/13346
;  PRIOR FILING DATE:  1997-07-22
;  NUMBER OF SEQ ID NOS:  30
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 17
;   LENGTH: 35100
;   TYPE: DNA
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-17

  Query Match             12.7%;  Score 127.6;  DB 4;  Length 35100;
  Best Local Similarity   66.9%;  Pred. No. 1.7e-17;
  Matches  198;  Conservative    0;  Mismatches   94;  Indels    4;  Gaps    1;

Qy        595 GCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 654
              | ||  |||  || |    ||||| ||||| ||||    ||||| ||||| ||||    |
Db      24611 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24552

Qy        655 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAG 714
              |||| ||||| ||||    ||||| ||||| ||||    |||||||||| ||||    |
Db      24551 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24492

Qy        715 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 774
              |||| ||||| ||||    |||||||||| ||||    ||||| ||||| ||||    |
Db      24491 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24432

Qy        775 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT----GCCTGCTGGCTT 830
              |||| ||||| ||||    ||||| ||||| ||||    |||||    |  ||||| |
Db      24431 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGG 24372

Qy        831 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTG 886
              |  | || |  ||||   |||||  || |  |||| |  ||||| ||| |  ||||
Db      24371 GGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTGCCGGGGCTCCTGGGGTGCTG 24316


RESULT 4
PCT-US93-06251-23/c
; Sequence 23, Application PC/TUS9306251
;  GENERAL INFORMATION:
;    APPLICANT:  Wickstrom, Eric and Rife, Jason P.
;    TITLE OF INVENTION:  Trivalent Synthesis of Oligonucleotides Containing
;    TITLE OF INVENTION:  Stereospecific Alkylphosphonates and Arylphosphonates
;    NUMBER OF SEQUENCES:  93
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCULLY, SCOTT, MURPHY & PRESSER
;      STREET:  400 Garden City Plaza
;      CITY:  Garden City
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  11530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/06251
;      FILING DATE:  19930630
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DiGiglio, Frank S.
;      REGISTRATION NUMBER:  31,346
;      REFERENCE/DOCKET NUMBER:  8586
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  516-742-4343
;      TELEFAX:  516-742-4366
;      TELEX:  230 901 SANS UR
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11558 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
PCT-US93-06251-23

  Query Match             11.0%;  Score 109.8;  DB 5;  Length 11558;
  Best Local Similarity   48.2%;  Pred. No. 5.9e-14;
```

```
 Matches  338;  Conservative    0;  Mismatches  362;  Indels    1;  Gaps    1;

Qy        252 GGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTG 311
              || ||  | ||| |||  ||   | |||   | | | ||   |||  |  ||   | | |
Db       6384 GGTGAACGAAGGACTGAATGTCAGATGGGTGGGTGGATGGATGGATGGATGGGTGGATGG 6325

Qy        312 CTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTG 371
              ||   |||  |  ||   | | | | | |||   | | | ||   |||  |  |||  |
Db       6324 ATGGATGGATTGATGGATGGATGGATGGATGGGTGGATGGGTGGATGGATGGATGGGTGG 6265

Qy        372 CTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 431
              |||||  | | |  ||   | | | |||   | | | | | | | |||   | | | | |
Db       6264 ATGGCTGGATGGATGGATGGGTGGATGGATGGATAGGTGGATGGATGGATAGGTGGATGG 6205

Qy        432 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 491
              | | |||   | | |||  | | | |||   | | | | | | | |||   | | | | |
Db       6204 ATGGATGGGTGGATGGCTGGATGGATGGATGGGTGGATGGGTGGATGGATGGATGGATGG 6145

Qy        492 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGA 551
              | | |||   | | | | | | | |||   | | | | | | | |||| | | ||||
Db       6144 ATGGATGGATGGATGGATGGATGGGTGGATGGGTGGATGGGTGGATGGCTGGTTGGCTGG 6085

Qy        552 GGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAG 611
              |||  |  |||  | | | | | |  | ||| ||   |   |  ||||| | ||| |
Db       6084 ATGGATGGATGGGTGGATGGGTGGGTGGGTGGATGGATGGATGGATGGATGGATGGATGG 6025

Qy        612 TAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAG 671
              |||  | |||   | | |  ||   | |||   | | |   ||  |  ||   | | |
Db       6024 ATGGGTGGATGGCTGGTTGGATGGATGGGTGGATGGATGGATAGGTGGACGGATGGATGG 5965

Qy        672 TAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAG 731
              ||  |||||  | | |  ||  | ||||  | | |  ||  | |||   | | |
Db       5964 GTGGATGGCTGGATGGATGGGTGGATTGCTGGATGGATTGGTGGATGGGTGGATGGGTGG 5905

Qy        732 TAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAG 791
              |||    |||| ||  || | ||  |   ||  |      || | | ||| |   |
Db       5904 ATGGGTGGATGGG-TGAAAGGATGGAAGGAAGGAGGGAGCATGGACGGATGGACATGTGA 5846

Qy        792 TAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGG 851
              ||||  | ||| | | |    ||   |   | ||| | |   |   | |        ||
Db       5845 ATGGGCAGATGGACAGATGAACGGTTGGAAGGATGGATGGAAAGAAGGATGAGAGAAAGG 5786

Qy        852 GCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGGGCTGCCGGCC 911
              |   |   |       ||    ||  ||| |||| ||| ||| ||  |     ||
Db       5785 AAAGAAGGAAGGAAGGAGAGGTGTATGGACAGATGGGTGAAGAGAATAAAGGAAGGGGAA 5726

Qy        912 AGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGG 952
              ||  |   |  ||   |  || |||  ||| |  | | ||
Db       5725 AGAAGGAAGGAAGATGGATGGATGGCTTGACAAACAGATGG 5685


RESULT 5
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
;   GENERAL INFORMATION:
;        APPLICANT: Fleischmann et al.
;        TITLE OF INVENTION: The Nucleotide sequence of
;                            the Haemophilus influenzae Rd Genome, Fragments
;                            Thereof, and Uses Thereof
;        NUMBER OF SEQUENCES: 1
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: MD
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3 1/2 inch diskette
;             COMPUTER: Dell Pentium
;             OPERATING SYSTEM: MS DOS v6.22
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/557,884
;             FILING DATE: 25-Apr-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/476,102
;             FILING DATE: JUN-5-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Michelle S. Marks
;             REGISTRATION NUMBER: 41,971
;             REFERENCE/DOCKET NUMBER: PB186P3
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 301-309-8504
;             TELEFAX: 301-309-8439
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1830121 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

  Query Match             7.7%;  Score 77.4;  DB 4;  Length 1830121;
  Best Local Similarity   84.5%;  Pred. No. 3.1e-07;
  Matches   87;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy         449 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 508
               ||||| ||| || |||||||||||| ||| || |||||||||||| ||| || |||||||
Db     1543129 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 1543188

Qy         509 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGA 551
               ||||| ||| || |||||||||||| ||| || ||||||| ||
Db     1543189 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCAGA 1543231


RESULT 6
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
;   GENERAL INFORMATION:
;        APPLICANT: Robert D. Fleischmann
;                   Mark D. Adams
;                   Owen White
;                   Hamilton O. Smith
;                   J. Craig Venter
;        TITLE OF INVENTION: The Nucleotide sequence of
;                            the Haemophilus influenzae Rd Genome, Fragments
;                            Thereof, and Uses Thereof
;        NUMBER OF SEQUENCES: 1
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville,
;             STATE: MD
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3 1/2 inch diskette
;             COMPUTER: Dell Pentium
;             OPERATING SYSTEM: MS DOS v6.22
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/643,990A
;             FILING DATE: 23-Aug-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/487,429
;             FILING DATE: 1995-06-07
;             APPLICATION NUMBER: 08/426,787
;             FILING DATE: 1995-04-21
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Kenley K. Hoover
```

```
;             REGISTRATION NUMBER: 40,302
;             REFERENCE/DOCKET NUMBER: PB186P1C1
;          TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 301-610-5790
;             TELEFAX: 310-309-8439
;    INFORMATION FOR SEQ ID NO: 1:
;          SEQUENCE CHARACTERISTICS:
;             LENGTH: 1830121 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;          SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

  Query Match             7.7%;  Score 77.4;  DB 4;  Length 1830121;
  Best Local Similarity   84.5%;  Pred. No. 3.1e-07;
  Matches   87;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy        449 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 508
              ||||| ||| || |||||||||||| ||| || |||||||||||| ||| || |||||||
Db    1543129 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 1543188

Qy        509 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGA 551
              ||||| ||| || |||||||||||| ||| || ||||||| ||
Db    1543189 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCAGA 1543231


RESULT 7
US-08-533-669A-7/c
; Sequence 7, Application US/08533669A
; Patent No. 5834592
;  GENERAL INFORMATION:
;    APPLICANT:  Corixa Corporation
;    TITLE OF INVENTION:  LEISHMANIA ANTIGENS FOR USE IN THE
;    TITLE OF INVENTION:  THERAPY AND DIAGNOSIS OF LEISHMANIASIS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/533,669A
;      FILING DATE:  22-SEP-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1771 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1698
US-08-533-669A-7

  Query Match             6.4%;  Score 64.2;  DB 2;  Length 1771;
  Best Local Similarity   48.7%;  Pred. No. 7.5e-05;
  Matches  320;  Conservative    0;  Mismatches  328;  Indels    9;  Gaps    5;

Qy        207 GACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 266
              | | ||  |   | |||    |  ||  ||| |||  | |||        |||  || |
Db        881 GCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGTGCATGACATCCAGGCGGGCC 822

Qy        267 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 326
              || | ||||   |||| |  |||  | |   ||  |  |||| | ||   ||| | | |
Db        821 GCCTCCTCGGCCGCCT-CCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGAC 763

Qy        327 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 386
              ||  || ||||||    | ||| ||       | | |   | || || |   ||    ||
Db        762 GCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAG 703

Qy        387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 446
                | || ||  | |  || ||| | ||||  ||||||   ||  ||  ||| | |  | |
Db        702 CTCCGCGCGGAGACGCGC-TGCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGT 644

Qy        447 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 506
              || ||    |  | | |   || | ||||  ||||||    |  ||  ||| | |  | |
Db        643 GCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGGCCTCGT 584

Qy        507 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT---GCTTGCTGAGGGGACAGTAGG 563
              || ||    |  | | |   || | ||||  ||||   |||    | || |||     |
Db        583 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTG 524

Qy        564 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 623
              || || |   ||   || |  ||   ||  | |      || ||       ||| |   |
Db        523 CCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGCATGGCCTCCAGGCGGGCCG 464

Qy        624 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 683
              || || |      || | |  | ||  | ||    ||| | || || |  || |||   |
Db        463 CCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-GCCTCCTCGAGGGCCTGGACG 405

Qy        684 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 743
                 ||  |   |||  |  |  ||  |    ||| |   | || || |     |||| |.
Db        404 CGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCGCCTCCTCGGCTTCCTCCAGC 345

Qy        744 GCTGC---TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 800
              | ||    ||  | |||||      |   |   ||  || || ||||| |      |||
Db        344 TCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGC 285

Qy        801 TGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 857
              |||  | ||  | || |||| || |   || | |      ||  |  |   ||||
Db        284 ATGGCCTCCAGGCGCGCCGCCTCCTCGGCAGCCTCCAGCTCCGCGCGGAGACGCGCT 228


RESULT 8
US-08-511-872-1/c
; Sequence 1, Application US/08511872
; Patent No. 5965142
;  GENERAL INFORMATION:
;    APPLICANT:  Dillon, Davin C.
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Day, Craig H.
;    TITLE OF INVENTION:  POLYPEPTIDES AND METHODS FOR THE
;    TITLE OF INVENTION:  DETECTION OF L. tropica INFECTION
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  94104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/511,872
;      FILING DATE:  04-AUG-1995
;      CLASSIFICATION:  424
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5965142tenburg, Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  210121.405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANDBERRY
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1771 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1698
US-08-511-872-1

  Query Match             6.4%;  Score 64.2;  DB 2;  Length 1771;
  Best Local Similarity   48.7%;  Pred. No. 7.5e-05;
  Matches  320;  Conservative    0;  Mismatches  328;  Indels    9;  Gaps    5;

Qy        207 GACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 266
              | | ||  |   | |||     |  ||  ||| |||  | |||         |||  || |
Db        881 GCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGTGCATGACATCCAGGCGGGCC 822

Qy        267 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 326
              || | ||||   |||| |  |||  | |   ||  |  |||| | ||   ||| | | |
Db        821 GCCTCCTCGGCCGCCT-CCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGAC 763

Qy        327 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 386
              ||  || ||||||    | ||| ||       | | |   | || || |   ||    ||
Db        762 GCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAG 703

Qy        387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 446
              | || ||  | |  || ||| | ||||  ||||||   ||  ||  ||| | |  | |
Db        702 CTCCGCGCGGAGACGCGC-TGCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGT 644

Qy        447 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 506
              || ||    |  | |     || | ||||  ||||||    |  ||  ||| | |  | |
Db        643 GCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGGCCTCGT 584

Qy        507 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT---GCTTGCTGAGGGGACAGTAGG 563
              || ||    |  | | |   || | ||||  ||||  |||    | || |||     |
Db        583 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTG 524

Qy        564 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 623
              || || |   ||   || | ||  ||  | |    || ||       ||| |   |
Db        523 CCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGCATGGCCTCCAGGCGGGCCG 464

Qy        624 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 683
              || || |     || | |  | ||  | ||    ||| | || || |  || |||   |
Db        463 CCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-GCCTCCTCGAGGGCCTGGACG 405

Qy        684 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 743
              ||  |   |||  |  |  ||  |    ||| |   | || || |     |||| |
Db        404 CGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCGCCTCCTCGGCTTCCTCCAGC 345

Qy        744 GCTGC---TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 800
              | ||    ||  | |||||      |   |   ||  || || ||||| |      |||
Db        344 TCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGC 285

Qy        801 TGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 857
              |||  | ||  | || |||| || |   || | |      ||  |  |   ||||
Db        284 ATGGCCTCCAGGCGCGCCGCCTCCTCGGCAGCCTCCAGCTCCGCGCGGAGACGCGCT 228


RESULT 9
US-09-183-861-7/c
; Sequence 7, Application US/09183861
; Patent No. 6365165
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Campos-Neto, Antonio
;    APPLICANT:  Webb, John R.
;    APPLICANT:  Dillon, Davin C.
;    APPLICANT:  Skeiky, Yasir A.W.
;    TITLE OF INVENTION:  LEISHMANIA ANTIGENS FOR USEIN THE THERAPY AND
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/183,861
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/022,765
;      FILING DATE:  12-FEB-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.420C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1771 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1698
US-09-183-861-7

  Query Match             6.4%;  Score 64.2;  DB 4;  Length 1771;
  Best Local Similarity   48.7%;  Pred. No. 7.5e-05;
  Matches  320;  Conservative    0;  Mismatches  328;  Indels    9;  Gaps    5;

Qy        207 GACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 266
              | | ||  |   | |||     |  ||  ||| |||  | |||         |||  || |
Db        881 GCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGTGCATGACATCCAGGCGGGCC 822

Qy        267 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 326
              || | ||||   |||| |  |||  | |   ||  |  |||| | ||   ||| | | |
Db        821 GCCTCCTCGGCCGCCT-CCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGAC 763

Qy        327 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 386
              ||  || ||||||    | ||| ||       | | |   | || || |   ||    ||
Db        762 GCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAG 703

Qy        387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 446
              | || ||  | |  || ||| | ||||  ||||||   ||  ||  ||| | |  | |
Db        702 CTCCGCGCGGAGACGCGC-TGCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGT 644

Qy        447 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 506
              || ||    |  | |     || | ||||  ||||||    |  ||  ||| | |  | |
Db        643 GCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGGCCTCGT 584

Qy        507 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT---GCTTGCTGAGGGGACAGTAGG 563
              || ||    |  | | |   || | ||||  ||||  |||    | || |||     |
Db        583 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTG 524

Qy        564 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 623
```

```
Db        523 CCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGCATGGCCTCCAGGCGGGCCG 464

Qy        624 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 683
Db        463 CCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-GCCTCCTCGAGGGCCTGGACG 405

Qy        684 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 743
Db        404 CGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCGCCTCCTCGGCTTCCTCCAGC 345

Qy        744 GCTGC---TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 800
Db        344 TCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGC 285

Qy        801 TGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 857
Db        284 ATGGCCTCCAGGCGCGCCGCCTCCTCGGCAGCCTCCAGCTCCGCGCGGAGACGCGCT 228


RESULT 10
US-09-022-765-7/c
; Sequence 7, Application US/09022765
; Patent No. 6375955
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Campos-Neto, Antonio
;    APPLICANT:  Webb, John R.
;    APPLICANT:  Dillon, Davin C.
;    APPLICANT:  Skeiky, Yasir A.W.
;    TITLE OF INVENTION:  LEISHMANIA ANTIGENS FOR USEIN THE THERAPY AND
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/022,765
;      FILING DATE:  12-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.420C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1771 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1698
US-09-022-765-7

  Query Match             6.4%;  Score 64.2;  DB 4;  Length 1771;
  Best Local Similarity   48.7%;  Pred. No. 7.5e-05;
  Matches  320;  Conservative    0;  Mismatches  328;  Indels    9;  Gaps    5;

Qy        207 GACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 266
Db        881 GCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGTGCATGACATCCAGGCGGGCC 822

Qy        267 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 326
Db        821 GCCTCCTCGGCCGCCT-CCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGAC 763

Qy        327 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 386
Db        762 GCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAG 703

Qy        387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 446
Db        702 CTCCGCGCGGAGACGCGC-TGCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGT 644

Qy        447 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 506
Db        643 GCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGGCCTCGT 584

Qy        507 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT---GCTTGCTGAGGGGACAGTAGG 563
Db        583 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTG 524

Qy        564 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 623
Db        523 CCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGCATGGCCTCCAGGCGGGCCG 464

Qy        624 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 683
Db        463 CCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-GCCTCCTCGAGGGCCTGGACG 405

Qy        684 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 743
Db        404 CGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCGCCTCCTCGGCTTCCTCCAGC 345

Qy        744 GCTGC---TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 800
Db        344 TCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGC 285

Qy        801 TGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 857
Db        284 ATGGCCTCCAGGCGCGCCGCCTCCTCGGCAGCCTCCAGCTCCGCGCGGAGACGCGCT 228


RESULT 11
US-09-551-974A-7/c
; Sequence 7, Application US/09551974A
; Patent No. 6500437
; GENERAL INFORMATION:
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Campos-Neto, Antonio
;  APPLICANT:  Webb, John R.
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Skeiky, Yasir A.W.
;  TITLE OF INVENTION:  LEISHMANIA ANTIGENS FOR USE IN THE
;  TITLE OF INVENTION:  THERAPY AND DIAGNOSIS OF LEISHMANIASIS
;  FILE REFERENCE:  210121.420C5
;  CURRENT APPLICATION NUMBER: US/09/551,974A
;  CURRENT FILING DATE:  2000-04-14
;  NUMBER OF SEQ ID NOS:  101
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 1771
;   TYPE: DNA
;   ORGANISM: Leishmania tropica
US-09-551-974A-7

  Query Match             6.4%;  Score 64.2;  DB 4;  Length 1771;
  Best Local Similarity   48.7%;  Pred. No. 7.5e-05;
  Matches  320;  Conservative    0;  Mismatches  328;  Indels    9;  Gaps    5;

Qy        207 GACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 266
Db        881 GCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGTGCATGACATCCAGGCGGGCC 822
```

```
Qy       267 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 326
Db       821 GCCTCCTCGGCCGCCT-CCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGAC 763

Qy       327 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 386
Db       762 GCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAG 703

Qy       387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 446
Db       702 CTCCGCGCGGAGACGCGC-TGCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGT 644

Qy       447 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 506
Db       643 GCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGGCCTCGT 584

Qy       507 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT---GCTTGCTGAGGGGACAGTAGG 563
Db       583 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTG 524

Qy       564 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 623
Db       523 CCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGCATGGCCTCCAGGCGGGCCG 464

Qy       624 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 683
Db       463 CCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-GCCTCCTCGAGGGCCTGGACG 405

Qy       684 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 743
Db       404 CGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCGCCTCCTCGGCTTCCTCCAGC 345

Qy       744 GCTGC---TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 800
Db       344 TCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGC 285

Qy       801 TGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 857
Db       284 ATGGCCTCCAGGCGCGCCGCCTCCTCGGCAGCCTCCAGCTCCGCGCGGAGACGCGCT 228
```

```
RESULT 12
US-08-476-102A-4/c
; Sequence 4, Application US/08476102A
; Patent No. 6355450
;  GENERAL INFORMATION:
;       APPLICANT: Fleischmann, et al.
;       TITLE OF INVENTION: Nucleotide Sequence of the Haemphilus
; Patent No. 6355450
;                          influenzae Rd Genome, Fragments Thereof, and Uses Thereof
;       NUMBER OF SEQUENCES: 9
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Ave.
;            CITY: Rockville
;            STATE: MD
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/476,102A
;            FILING DATE: 07-Jun-1995
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Marks, Michelle S.
;            REGISTRATION NUMBER: 41,971
;            REFERENCE/DOCKET NUMBER: PB186P2
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 301-309-8504
;            TELEX: 301-309-8439
;  INFORMATION FOR SEQ ID NO: 4:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 3252 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-08-476-102A-4

  Query Match             5.8%;  Score 58;  DB 4;  Length 3252;
  Best Local Similarity   58.8%;  Pred. No. 0.0014;
  Matches  100;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy       405 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 464
Db       222 TTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGG 163

Qy       465 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 524
Db       162 TTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGG 103

Qy       525 TTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 574
Db       102 TTGGTTGGTTGGTTGGTTGGTTGGTTGGGCTGCATAGACTGTGCTTGCAT 53


RESULT 13
PCT-US96-05320A-714/c
; Sequence 714, Application PC/TUS9605320A
;  GENERAL INFORMATION:
;    APPLICANT:  Human Genome Sciences
;    APPLICANT:  9410 Key West Avenue
;    APPLICANT:  Rockville, MD 20850
;    APPLICANT:  United States of America
;    APPLICANT:  Johns Hopkins University
;    APPLICANT:  720 Rutland Avenue
;    APPLICANT:  Baltimore, MD 21205
;    APPLICANT:  United States of America
;    APPLICANT:  Mark D. Adams
;    APPLICANT:  Owen White
;    APPLICANT:  Hamilton O. Smith
;    APPLICANT:  J. Craig Venter
;    TITLE OF INVENTION:  Nucleotide Sequence of the Haemophilus Influenzae Rd Genome
;    NUMBER OF SEQUENCES:  48
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20003-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/05320A
;      FILING DATE:  April22, 1996
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/476,102
;      FILING DATE:  June 7, 1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/487,429
;      FILING DATE:  June 7, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Eric K. Steffe
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.014PC01
;    TELECOMMUNICATION INFORMATION:
```

```
;       TELEPHONE:  (202) 371-2600
;       TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  714:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  3252 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
PCT-US96-05320A-714

  Query Match             5.8%;  Score 58;  DB 5;  Length 3252;
  Best Local Similarity   58.8%;  Pred. No. 0.0014;
  Matches  100;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy        405 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 464
              ||| ||| | | | |  || ||| ||| | | | |  || ||| ||| | | | |  ||
Db        222 TTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGG 163

Qy        465 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 524
              ||| ||| | | | |  || ||| ||| | | | |  || ||| ||| | | | |  ||
Db        162 TTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGG 103

Qy        525 TTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 574
              ||| ||| | | | |  || ||| |  ||   ||       |||||||
Db        102 TTGGTTGGTTGGTTGGTTGGTTGGTTGGGCTGCATAGACTGTGCTTGCAT 53


RESULT 14
US-09-146-053-6/c
; Sequence 6, Application US/09146053A
; Patent No. 6399349
; GENERAL INFORMATION:
;  APPLICANT: Ryan, James W.
;  APPLICANT: Sprinkle, Terry Joe Curtis
;  APPLICANT: Venema, Richard C.
;  TITLE OF INVENTION: Human Aminopeptidase P Gene
;  FILE REFERENCE: MCG103
;  CURRENT APPLICATION NUMBER: US/09/146,053A
;  CURRENT FILING DATE: 1998-09-02
;  EARLIER APPLICATION NUMBER: 60/057,854
;  EARLIER FILING DATE: 1997-09-02
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 45546
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-146-053-6

  Query Match             5.6%;  Score 56;  DB 4;  Length 45546;
  Best Local Similarity   59.4%;  Pred. No. 0.0045;
  Matches   95;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy        390 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 449
              | ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||
Db      18742 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 18683

Qy        450 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 509
              ||| ||  |   ||  ||| ||| ||  ||  ||| ||| ||  ||  | | ||| ||
Db      18682 CTCTCTCTCTCTCTCTCTCTTTCTCTCTCCCTCTTTCTTTCTTTCTTTCTTTCTTTCT 18623

Qy        510 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              | ||  ||  | | ||| ||| ||  ||  | | ||| ||
Db      18622 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTCCT 18583


RESULT 15
US-08-595-937A-1/c
; Sequence 1, Application US/08595937A
; Patent No. 5919919
; Patent No. 5919919 5786467
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  A Brassica sp. Gene Promoter Highly Expressed
;    TITLE OF INVENTION:  During Tapetum Development
;    NUMBER OF SEQUENCES:  2
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/595,937A
;      FILING DATE:
;      CLASSIFICATION:  800
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1960 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-595-937A-1

  Query Match             5.6%;  Score 55.6;  DB 2;  Length 1960;
  Best Local Similarity   45.6%;  Pred. No. 0.0041;
  Matches  277;  Conservative    0;  Mismatches  324;  Indels    7;  Gaps    2;

Qy        265 CTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGT 324
              | | || |  |  ||  || ||||||      |||  ||  |  |||| |      ||||
Db        812 CGGTTTTCCAGTCGGTTTGTTTGCTGCCTCTCCAGCCGGTTTATTTGCAGCCCCTCCAGT 753

Qy        325 AGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGT 384
              ||  |  ||| | | |   ||| | |       ||   ||    ||  ||| ||
Db        752 CGGTTTATTTGTTGGTTTATCTGTTGGTCCTCCGGCAGGTTTATCTTTTGGTTTATCTAC 693

Qy        385 AGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 444
              |  |  |  ||||| |   || | |  | |  ||  | ||   |  | |    ||
Db        692 TGATCCTCCAGCTGGTTTATCTGTTGGTCCACCAGCGGGTTTATCTTTTGGTTTATTTGA 633

Qy        445 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTC----GCTGGCCTGCTTGCTTGCTCGCTGGC 500
              |  ||  |  ||||| | ||| || |||     | |||  |  |  ||  | |
Db        632 TGCCTCTCCAGCTGGTTTACTTTCTGGCTTATCTGTTGGTCCACCAGCGAGTTTATCTTT 573

Qy        501 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 560
              | ||   || |   | || || | || ||  ||||| |   || ||      |||
Db        572 TGGTTTATGTGATGTTTCTCCAGCCGGTTTACTTTCTGGCTTATCTGTTGGTCCACCAGC 513

Qy        561 AGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGC 620
              ||  |   |  | | | | | || | | | || ||  |  |  |||  | |      ||
Db        512 GGGTTTATCTTTTGGTTTATGTGATGTTTCTCCAGCTGGTTTACTTTCTGGCTTATTTGT 453

Qy        621 TGGGCTGCTAGTAGGGCT---GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 677
              |||  |  |  |   |     | || |  |   | |  ||     |||  || | |   |
Db        452 TGGTCCTCCAACGGATTTATCTTTTGGTTTATTTGATGTCTCTCCAGCTTCTGGCATATC 393

Qy        678 TGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 737
              || ||  |  | ||  ||  |         |  |  |    ||   |  | |||
Db        392 TGTTGATCCACCAGCGGGTTTATCTTTTGGTTTATTTGATGCCTCTCCAACCAGTTTACT 333

Qy        738 TCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGC 797
              | |||   |    || ||  |  |  |  |  |    | | |    | || |    |
Db        332 TTCTGTCTTATCTGTTGGTCCTCCAGCGGGTTTATCAACCGGTTTACTTCCTGGCTTATC 273

Qy        798 TGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 857
              || ||| |  | ||  ||  |  ||||||||||| ||| || | ||    || || || |
Db        272 TGTTGGTCCTCCAGCGGGTTTATCTGCTGGCTTACTTTCTGGTTTATCCGTTGGTCCTCC 213

Qy        858 TGCCTGCT 865
              ||  | |
Db        212 AGCGGGTT 205


RESULT 16
US-09-539-333D-1/c
; Sequence 1, Application US/09539333D
```

```
; Patent No. 6476208
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Chumakov, Ilya
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Bihain, Bernard
;  APPLICANT:  Essioux, Laurent
;  TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
;  FILE REFERENCE: GENSET.047AUS
;  CURRENT APPLICATION NUMBER: US/09/539,333D
;  CURRENT FILING DATE:  2000-03-30
;  PRIOR APPLICATION NUMBER: US 60/126,903
;  PRIOR FILING DATE: 1999-03-30
;  PRIOR APPLICATION NUMBER: US 60/131,971
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: US 60/132,065
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: US 60/143,928
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 60/145,915
;  PRIOR FILING DATE: 1999-07-27
;  PRIOR APPLICATION NUMBER: US 60/146,453
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 60/146,452
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 60/162,288
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: US 09/416,384
;  PRIOR FILING DATE: 1999-10-12
;  NUMBER OF SEQ ID NOS: 231
;  SOFTWARE: Patent.pm
; SEQ ID NO 1
;   LENGTH: 319608
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 31..1107
;   OTHER INFORMATION: 5'regulatory region g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 1108..1289
;   OTHER INFORMATION: exon A g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 14877..14920
;   OTHER INFORMATION: exon B g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 18778..18862
;   OTHER INFORMATION: exon Bbis g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 25593..25740
;   OTHER INFORMATION: exon C g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 29388..29502
;   OTHER INFORMATION: exon D g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 29967..30282
;   OTHER INFORMATION: exon E g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 64666..64812
;   OTHER INFORMATION: exon F g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 65505..65853
;   OTHER INFORMATION: exon G g35018 gene
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 65854..67854
;   OTHER INFORMATION: 3'regulatory region g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 94124..94964
;   OTHER INFORMATION: exon g35017
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 201188..201234
;   OTHER INFORMATION: exon S g35030 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 214676..214793
;   OTHER INFORMATION: exon T g35030 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 215702..215746
;   OTHER INFORMATION: exon U g35030 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 216836..216915
;   OTHER INFORMATION: exon V g35030 gene
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 213818..215818
;   OTHER INFORMATION: 3'regulatory region g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 215819..215941
;   OTHER INFORMATION: exon R complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 215819..215975
;   OTHER INFORMATION: exon Rbis complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 216661..216952
;   OTHER INFORMATION: exon Qbis complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 216661..217061
;   OTHER INFORMATION: exon Q complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 217027..217061
;   OTHER INFORMATION: exon Q1 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 229647..229742
;   OTHER INFORMATION: exon X complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 230408..230721
;   OTHER INFORMATION: exon P complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 231272..231412
;   OTHER INFORMATION: exon Obis complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 231787..231880
;   OTHER INFORMATION: exon O2 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 231870..231879
;   OTHER INFORMATION: exon O1 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 234174..234321
;   OTHER INFORMATION: exon O complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
```

```
;   LOCATION: 237406..237428
;   OTHER INFORMATION: exon Nbis complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 239719..239807
;   OTHER INFORMATION: exon N2 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 239719..239853
;   OTHER INFORMATION: exon N complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240569
;   OTHER INFORMATION: exon M1117 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240596
;   OTHER INFORMATION: exon M1090 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240617
;   OTHER INFORMATION: exon M1069 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240644
;   OTHER INFORMATION: exon MS2 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240824
;   OTHER INFORMATION: exon M862 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240994
;   OTHER INFORMATION: exon M692 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..241685
;   OTHER INFORMATION: exon M1 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240800..240993
;   OTHER INFORMATION: exon MS1 complement g34872 gene
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 241686..243685
;   OTHER INFORMATION: 5'regulatory region g34872 gene
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 290652..292652
;   OTHER INFORMATION: 3'regulatory region g34665 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 292653..292841

  Query Match             5.3%;  Score 52.6;  DB 4;  Length 319608;
  Best Local Similarity   52.5%;  Pred. No. 0.026;
  Matches  115;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy        328 CCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGG 387
              ||| | | |   ||  ||| | | ||       |       || ||  ||| ||  |
Db      61813 CCTTCCTTCCATCTTTCCTTCCTTCTTTCCTTCCTTCCTTCCTTCTTTCTTCCTTCTTCC 61754

Qy        388 GCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 447
              | ||| ||  | | ||| ||| ||  ||  |||   |  || ||  ||    | |||
Db      61753 CTTTCTCCCTCCCTTCCTTCCTTTCTTCCTTACCTCAATCTTTCCTTCCTTCTTTCCTTC 61694

Qy        448 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 507
              ||| ||  ||  | |  || ||| ||  ||  | | | |  || ||  ||  | | |||
Db      61693 CTTCCTTCCTTCCTTCATTCCTTCCTTCCTTCCTTCCATCTTTCCTTCCTTCCTTCCTTC 61634

Qy        508 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 546
              ||| ||  ||  | | ||| ||| ||  ||  | | |||
Db      61633 CTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTT 61595
```

```
RESULT 17
US-09-679-409-1/c
; Sequence 1, Application US/09679409
; Patent No. 6555316
; GENERAL INFORMATION:
;  APPLICANT:  Cohen, Daniel
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Chumakov, Ilya
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Essioux, Laurent
;  TITLE OF INVENTION:  SCHIZOPHRENIA ASSOCIATED GENE, PROTEINS AND BIALLELIC MARKERS
;  FILE REFERENCE:  53.US15.CIP
;  CURRENT APPLICATION NUMBER: US/09/679,409
;  CURRENT FILING DATE:  2000-10-03
;  PRIOR APPLICATION NUMBER:  09/539,333
;  PRIOR FILING DATE:  2000-03-03
;  PRIOR APPLICATION NUMBER:  09/416,384
;  PRIOR FILING DATE:  1999-10-12
;  PRIOR APPLICATION NUMBER:  60/168,088
;  PRIOR FILING DATE:  1999-11-30
;  NUMBER OF SEQ ID NOS:  134
;  SOFTWARE:  Patent.pm
; SEQ ID NO 1
;   LENGTH: 319608
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 199122..201122
;   OTHER INFORMATION: 5'regulatory region
;   NAME/KEY: exon
;   LOCATION: 201123..201234
;   OTHER INFORMATION: exon S
;   NAME/KEY: exon
;   LOCATION: 201123..201560
;   OTHER INFORMATION: exon S2
;   NAME/KEY: exon
;   LOCATION: 214676..214793
;   OTHER INFORMATION: exon T
;   NAME/KEY: exon
;   LOCATION: 215702..215746
;   OTHER INFORMATION: exon U
;   NAME/KEY: exon
;   LOCATION: 216836..216994
;   OTHER INFORMATION: exon V
;   NAME/KEY: exon
;   LOCATION: 216836..217077
;   OTHER INFORMATION: exon V2
;   NAME/KEY: exon
;   LOCATION: 217671..217764
;   OTHER INFORMATION: exon V1
;   NAME/KEY: exon
;   LOCATION: 227655..227736
;   OTHER INFORMATION: exon V4
;   NAME/KEY: exon
;   LOCATION: 238715..238919
;   OTHER INFORMATION: exon V3
;   NAME/KEY: exon
;   LOCATION: 240440..240673
;   OTHER INFORMATION: exon W
;   NAME/KEY: exon
;   LOCATION: 240440..241153
;   OTHER INFORMATION: exon W2
;   NAME/KEY: exon
;   LOCATION: 241072..241291
;   OTHER INFORMATION: exon X
;   NAME/KEY: exon
;   LOCATION: 244353..244561
;   OTHER INFORMATION: exon Y
;   NAME/KEY: exon
;   LOCATION: 246273..247802
```

```
;  OTHER INFORMATION: exon 2
;  NAME/KEY: misc_feature
;  LOCATION: 247803..249803
;  OTHER INFORMATION: 3'regulatory region
;  NAME/KEY: allele
;  LOCATION: 8316
;  OTHER INFORMATION: 99-27943-150  : polymorphic base G or C
;  NAME/KEY: allele
;  LOCATION: 21672
;  OTHER INFORMATION: 99-27935-193  : polymorphic base G or C
;  NAME/KEY: allele
;  LOCATION: 65485
;  OTHER INFORMATION: 8-128-33  : polymorphic base C or T
;  NAME/KEY: allele
;  LOCATION: 95396
;  OTHER INFORMATION: 99-31960-363  : polymorphic base A or G
;  NAME/KEY: allele
;  LOCATION: 107281
;  OTHER INFORMATION: 99-24656-260  : polymorphic base A or G
;  NAME/KEY: allele
;  LOCATION: 160640
;  OTHER INFORMATION: 99-24639-163  : polymorphic base A or C
;  NAME/KEY: allele
;  LOCATION: 160876
;  OTHER INFORMATION: 99-24634-108  : polymorphic base A or T
;  NAME/KEY: allele
;  LOCATION: 168974
;  OTHER INFORMATION: 99-7652-162  : polymorphic base A or G
;  NAME/KEY: allele
;  LOCATION: 170810
;  OTHER INFORMATION: 99-16100-147  : polymorphic base A or G
;  NAME/KEY: allele
;  LOCATION: 173358
;  OTHER INFORMATION: 99-5862-167  : polymorphic base A or G
;  NAME/KEY: allele
;  LOCATION: 189957
;  OTHER INFORMATION: 99-5919-215  : polymorphic base A or G
;  NAME/KEY: allele
;  LOCATION: 197163
;  OTHER INFORMATION: 99-24658-410  : polymorphic base A or G
;  NAME/KEY: allele
;  LOCATION: 200778
;  OTHER INFORMATION: 8-303-235  : polymorphic base A or G
;  NAME/KEY: allele
;  LOCATION: 202651
;  OTHER INFORMATION: 8-300-221  : polymorphic base A or G
;  NAME/KEY: allele
;  LOCATION: 202679
;  OTHER INFORMATION: 8-300-193  : polymorphic base A or G
;  NAME/KEY: allele
;  LOCATION: 203378
;  OTHER INFORMATION: 8-299-128  : polymorphic base A or T
;  NAME/KEY: allele
;  LOCATION: 204138
;  OTHER INFORMATION: 8-296-213  : polymorphic base A or T
;  NAME/KEY: allele
;  LOCATION: 204605
;  OTHER INFORMATION: 8-252-190  : polymorphic base C or T
;  NAME/KEY: allele
;  LOCATION: 204934
;  OTHER INFORMATION: 99-24644-194  : polymorphic base A or G
;  NAME/KEY: allele
;  LOCATION: 205206
;  OTHER INFORMATION: 8-295-248  : polymorphic base A or C
;  NAME/KEY: allele
;  LOCATION: 205329
;  OTHER INFORMATION: 8-295-125  : polymorphic base C or T
;  NAME/KEY: allele
;  LOCATION: 206064
;  OTHER INFORMATION: 8-293-130  : polymorphic base A or G
;  NAME/KEY: allele
;  LOCATION: 206545
;  OTHER INFORMATION: 8-292-198  : polymorphic base A or G
;  NAME/KEY: allele
;  LOCATION: 207313
;  OTHER INFORMATION: 8-251-322  : polymorphic base A or G
;  NAME/KEY: allele
;  LOCATION: 208285
;  OTHER INFORMATION: 8-289-322  : polymorphic base A or G
;  NAME/KEY: allele
;  LOCATION: 208960
;  OTHER INFORMATION: 8-287-249  : polymorphic base C or T
;  NAME/KEY: allele
;  LOCATION: 209123
;  OTHER INFORMATION: 8-287-86  : polymorphic base A or T
;  NAME/KEY: allele
;  LOCATION: 209631
;  OTHER INFORMATION: 8-285-319  : polymorphic base A or G
;  NAME/KEY: allele
;  LOCATION: 210361
;  OTHER INFORMATION: 8-283-278  : polymorphic base G or C
;  NAME/KEY: allele
;  LOCATION: 210463
;  OTHER INFORMATION: 8-283-176  : polymorphic base A or G
;  NAME/KEY: allele
;  LOCATION: 210486
;  OTHER INFORMATION: 8-283-153  : polymorphic base G or C
;  NAME/KEY: allele
;  LOCATION: 210583
;  OTHER INFORMATION: 8-283-56  : polymorphic base C or T
;  NAME/KEY: allele
;  LOCATION: 210879
;  OTHER INFORMATION: 8-282-345  : polymorphic base G or C
;  NAME/KEY: allele
;  LOCATION: 210964
;  OTHER INFORMATION: 8-282-260  : polymorphic base G or T
;  NAME/KEY: allele
;  LOCATION: 210979
;  OTHER INFORMATION: 8-282-245  : polymorphic base A or C
;  NAME/KEY: allele
;  LOCATION: 211050
;  OTHER INFORMATION: 8-282-174  : variable motif AAAGG or GAAGGAAGGAAGGAAGGAAGA
;  NAME/KEY: allele
;  LOCATION: 211132
;  OTHER INFORMATION: 8-282-92  : polymorphic base A or T
;  NAME/KEY: allele
;  LOCATION: 211247
;  OTHER INFORMATION: 8-281-367  : polymorphic base A or G
;  NAME/KEY: allele
;  LOCATION: 211315
;  OTHER INFORMATION: 8-281-299  : polymorphic base A or G
;  NAME/KEY: allele
;  LOCATION: 211366
;  OTHER INFORMATION: 8-281-248  : polymorphic base G or C
;  NAME/KEY: allele
;  LOCATION: 212520
;  OTHER INFORMATION: 8-279-197  : polymorphic base A or C
;  NAME/KEY: allele
;  LOCATION: 212821
;  OTHER INFORMATION: 8-278-289  : polymorphic base C or T
;  NAME/KEY: allele

  Query Match             5.3%;  Score 52.6;  DB 4;  Length 319608;
  Best Local Similarity   52.5%;  Pred. No. 0.026;
  Matches  115;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy        328 CCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGG 387
              ||| | | |   ||  ||| | | ||       |       || ||  ||| ||  |
Db      61813 CCTTCCTTCCATCTTTCCTTCCTTCTTTCCTTCCTTCCTTCCTTCTTTCTTCCTTCTTCC 61754

Qy        388 GCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 447
                | ||| ||  | | ||| ||| ||  ||  |||   |  || ||  ||    | |||
Db      61753 CTTTCTCCCTCCCTTCCTTCCTTTCTTCCTTACCTCAATCTTTCCTTCCTTCTTTCCTTC 61694

Qy        448 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 507
              ||| ||  ||  | |  || ||| ||  ||  | | | |  || ||  ||  | | |||
```

```
Db      61693 CTTCCTTCCTTCCTTCATTCCTTCCTTCCTTCCTTCCATCTTTCCTTCCTTCCTTCCTTC 61634

Qy        508 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 546
              ||| ||  ||  | | ||| ||| ||  ||  | | |||
Db      61633 CTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTT 61595


RESULT 18
US-08-332-766A-20/c
; Sequence 20, Application US/08332766A
; Patent No. 5843647
;  GENERAL INFORMATION:
;    APPLICANT:  JEFFREYS, Alec J.
;    APPLICANT:  ARMOUR, John
;    TITLE OF INVENTION:  SIMPLE TANDEM REPEATS
;    NUMBER OF SEQUENCES:  125
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D. C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/332,766A
;      FILING DATE:  01-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9326052.9
;      FILING DATE:  21-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BIRD, Donald J.
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  217211/M94/0434/GB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  350 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-332-766A-20

  Query Match             5.2%;  Score 52.4;  DB 2;  Length 350;
  Best Local Similarity   54.5%;  Pred. No. 0.015;
  Matches  104;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy        360 ACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 419
              ||| | |     ||| ||  ||    |     ||  ||| | | ||| ||| ||| |||
Db        334 ACATTTGCAAGCCTGCCTGCCTGCCTGCCTGCCTGCCTTCCTTCCTTNCTCCCTGA 275

Qy        420 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 479
              |   ||| ||| ||  ||     | ||| ||| ||  ||  |   ||| ||  ||  ||
Db        274 CTGCCTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCCTTCCTACCTTCCTACCTTCCTTC 215

Qy        480 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 539
              | |  || ||| ||  ||     | ||| ||| ||   |  | | ||| ||| ||  |||
Db        214 CTTCTTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCCTGCCTGC 155

Qy        540 CCTGCTTGCTG 550
              |   ||  |||
Db        154 CTGCCTGCCTG 144


RESULT 19
US-08-284-941-1
; Sequence 1, Application US/08284941
; Patent No. 5863756
;  GENERAL INFORMATION:
;    APPLICANT:  BARR, PHILIP J
;    APPLICANT:  KIEFER, MICHAEL C
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR PACE 4 AND
;    TITLE OF INVENTION:  PACE 4.1 GENE AND POLYPEPTIDES IN CELLS
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  COOLEY GODWARD CASTRO HUDDLESON & TATUM
;      STREET:  FIVE PALO ALTO SQUARE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/284,941
;      FILING DATE:  2 August 1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  NEELEY PH.D., RICHARD L.
;      REGISTRATION NUMBER:  30092
;      REFERENCE/DOCKET NUMBER:  CHIR-009/01US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 843-5070
;      TELEFAX:  (415) 857-0663
;      TELEX:  380816 COOLEY PA
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4403 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  170..3077
US-08-284-941-1

  Query Match             5.2%;  Score 51.8;  DB 2;  Length 4403;
  Best Local Similarity   57.9%;  Pred. No. 0.025;
  Matches   92;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy         20 CCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCTCCGCGCCGCCGGGAG 79
              || ||| ||  ||| |  | |   |||||| |       || | | |   ||| ||| | |
Db        123 CCCCCGGCGCACTCGCAGCTCGGGCGCCGCGCGAGCCTGTCGCCGCTATGCCTCCGCGCG 182

Qy         80 GGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGACCGAGGGCCCCG 139
                |  ||| | ||||||| | |  | || |  || |||    | ||| |||||   | |||
Db        183 CGCCGCCTGCGCCCGGGCCCCGGCCGCCGCCCCGGGCCGCCGCCGCCACCGACACCGCCG 242

Qy        140 GGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGG 178
              | |  ||||||    ||    |||| |||| || || |
Db        243 CGGGCGCGGGGGGCGCGGGGGGCGCGGGGGGCGCCGGCG 281


RESULT 20
US-08-447-642-1
; Sequence 1, Application US/08447642
; Patent No. 5989890
;  GENERAL INFORMATION:
;    APPLICANT:  BARR, PHILIP J
;    APPLICANT:  KIEFER, MICHAEL C
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR PACE 4 AND
;    TITLE OF INVENTION:  PACE 4.1 GENE AND POLYPEPTIDES IN CELLS
;    NUMBER OF SEQUENCES:  16
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  COOLEY GODWARD CASTRO HUDDLESON & TATUM
;      STREET:  FIVE PALO ALTO SQUARE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/447,642
;      FILING DATE:  23-MAY-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/284,941
;      FILING DATE:  2 August 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  NEELEY PH.D., RICHARD L.
;      REGISTRATION NUMBER:  30092
;      REFERENCE/DOCKET NUMBER:  CHIR-009/01US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 843-5070
;      TELEFAX:  (415) 857-0663
;      TELEX:  380816 COOLEY PA
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4403 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  170..3077
US-08-447-642-1

  Query Match             5.2%;  Score 51.8;  DB 2;  Length 4403;
  Best Local Similarity   57.9%;  Pred. No. 0.025;
  Matches   92;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy         20 CCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCTCCGCGCCGCCGGGAG 79
              || ||| ||  ||| |  | |   |||||| |      || | | |   ||| ||| | |
Db        123 CCCCCGGCGCACTCGCAGCTCGGGCGCCGCGCGAGCCTGTCGCCGCTATGCCTCCGCGCG 182

Qy         80 GGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGACCGAGGGCCCCG 139
               |  ||| | |||||||| | |  | || |  || |||     | ||| |||||   | |||
Db        183 CGCCGCCTGCGCCCGGGCCCCGGCCGCCGCCCCGGGCCGCCGCCACCGACACCGCCG 242

Qy        140 GGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGG 178
              | |  ||||||    ||    |||| |||| || || |
Db        243 CGGGCGCGGGGGGCGCGGGGGGCGCGGGGGGCGCCGGCG 281


RESULT 21
US-09-236-503-1
; Sequence 1, Application US/09236503
; Patent No. 6277590
; GENERAL INFORMATION:
;  APPLICANT: Barr, Philip J
;  APPLICANT: Kiefer, Michael C
;  TITLE OF INVENTION: Compositions and Methods for PACE 4 and 4.1 Gene and
;  TITLE OF INVENTION: Polypeptides in Cells
;  FILE REFERENCE: CHIR-009/04US
;  CURRENT APPLICATION NUMBER: US/09/236,503
;  CURRENT FILING DATE: 1999-01-25
;  EARLIER APPLICATION NUMBER: 08/447,642
;  EARLIER FILING DATE: 1995-05-23
;  EARLIER APPLICATION NUMBER: 08/284,941
;  EARLIER FILING DATE: 1994-08-02
;  EARLIER APPLICATION NUMBER: 07/848,629
;  EARLIER FILING DATE: 1992-03-09
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 4403
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (170)..(3076)
US-09-236-503-1

  Query Match             5.2%;  Score 51.8;  DB 3;  Length 4403;
  Best Local Similarity   57.9%;  Pred. No. 0.025;
  Matches   92;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy         20 CCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCTCCGCGCCGCCGGGAG 79
              || ||| ||  ||| |  | |   |||||| |      || | | |   ||| ||| | |
Db        123 CCCCCGGCGCACTCGCAGCTCGGGCGCCGCGCGAGCCTGTCGCCGCTATGCCTCCGCGCG 182

Qy         80 GGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGACCGAGGGCCCCG 139
               |  ||| | |||||||| | |  | || |  || |||     | ||| |||||   | |||
Db        183 CGCCGCCTGCGCCCGGGCCCCGGCCGCCGCCCCGGGCCGCCGCCACCGACACCGCCG 242

Qy        140 GGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGG 178
              | |  ||||||    ||    |||| |||| || || |
Db        243 CGGGCGCGGGGGGCGCGGGGGGCGCGGGGGGCGCCGGCG 281


RESULT 22
PCT-US93-02147A-1
; Sequence 1, Application PC/TUS9302147A
;  GENERAL INFORMATION:
;    APPLICANT:  BARR, PHILIP J
;    APPLICANT:  KIEFER, MICHAEL C
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR PACE 4 AND
;    TITLE OF INVENTION:  PACE.4.1 GENE AND POLYPEPTIDES IN CELLS
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  COOLEY GODWARD CASTRO HUDDLESON & TATUM
;      STREET:  FIVE PALO ALTO SQUARE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/02147A
;      FILING DATE:  19930309
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/848,629
;      FILING DATE:  09-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  NEELEY PH.D., RICHARD L
;      REGISTRATION NUMBER:  30092
;      REFERENCE/DOCKET NUMBER:  CHIR-009/00US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-7622
;      TELEFAX:  (415) 857-0663
;      TELEX:  380816 COOLEY PA
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4403 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (cDNA)
;    FEATURE:
```

```
;       NAME/KEY:  CDS
;       LOCATION:  170..3077
PCT-US93-02147A-1

  Query Match             5.2%;  Score 51.8;  DB 5;  Length 4403;
  Best Local Similarity   57.9%;  Pred. No. 0.025;
  Matches   92;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy         20 CCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCTCCGCGCCGCCGGGAG 79
              || ||| ||  ||| |  | |   |||||| |      || | | |   ||| ||| | |
Db        123 CCCCCGGCGCACTCGCAGCTCGGGCGCCGCGCGAGCCTGTCGCCGCTATGCCTCCGCGCG 182

Qy         80 GGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGACCGAGGGCCCCG 139
               |  ||| | ||||||| | |  | || |  || |||    | ||| |||||   | |||
Db        183 CGCCGCCTGCGCCCGGGCCCCGGCCGCCGCCCCGGGCCGCCGCCGCCACCGACACCGCCG 242

Qy        140 GGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGG 178
              | |  ||||||    ||    |||| |||| || || |
Db        243 CGGGCGCGGGGGGCGCGGGGGGCGCGGGGGGCGCCGGCG 281


RESULT 23
US-08-332-766A-24/c
; Sequence 24, Application US/08332766A
; Patent No. 5843647
;  GENERAL INFORMATION:
;    APPLICANT:  JEFFREYS, Alec J.
;    APPLICANT:  ARMOUR, John
;    TITLE OF INVENTION:  SIMPLE TANDEM REPEATS
;    NUMBER OF SEQUENCES:  125
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D. C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/332,766A
;      FILING DATE:  01-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9326052.9
;      FILING DATE:  21-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BIRD, Donald J.
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  217211/M94/0434/GB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  538 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-332-766A-24

  Query Match             5.2%;  Score 51.6;  DB 2;  Length 538;
  Best Local Similarity   56.5%;  Pred. No. 0.023;
  Matches   96;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        380 CTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 439
              |||||       ||| ||  |   ||  ||  ||| ||  |   ||  ||  ||| ||
Db        299 CTAGTCCTTTCCCTCCCTCCCTCCCTCCCTCCCTCCCTTCCTCCCTCCCTCCCTCCCTCC 240

Qy        440 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 499
              | | ||  ||| ||  ||  ||| ||| ||| |   ||  ||| ||| ||| |   ||
Db        239 CTTCCTCCCTTCCTTCCTTCCCTCCTTCCTTCCCTCCTTCCCTCCTTCCTTCCCTCCTTC 180

Qy        500 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              | | | | ||| ||  ||  ||| ||| ||| ||  ||  | | ||| ||
Db        179 CTTCCCTCCTTCCTTCCTTCCCTCCTTCCTTCCTTCCTTCCATTCTTTCT 130


RESULT 24
US-09-268-992-7/c
; Sequence 7, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
;  APPLICANT: Chen, H.
;  APPLICANT: Freimer, N.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
;  TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
;  FILE REFERENCE: 7853-138
;  CURRENT APPLICATION NUMBER: US/09/268,992
;  CURRENT FILING DATE: 1999-03-16
;  EARLIER APPLICATION NUMBER: 09/236,134
;  EARLIER FILING DATE: 1999-01-22
;  EARLIER APPLICATION NUMBER: 60/106,056
;  EARLIER FILING DATE: 1998-10-28
;  EARLIER APPLICATION NUMBER: 60/088,312
;  EARLIER FILING DATE: 1998-06-05
;  EARLIER APPLICATION NUMBER: 60/078,044
;  EARLIER FILING DATE: 1998-03-16
;  NUMBER OF SEQ ID NOS: 84
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 72604
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: all n positions
;   OTHER INFORMATION: n=a, c, g, or t
US-09-268-992-7

  Query Match             5.1%;  Score 51.4;  DB 4;  Length 72604;
  Best Local Similarity   58.0%;  Pred. No. 0.04;
  Matches   91;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy        393 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 452
              ||  |  | | |||  || ||| ||  | | ||| ||| ||  ||  | | ||| ||| |
Db      45313 TCTTTCTCTTTCTTCTTTTCTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTYCTTYC 45254

Qy        453 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 512
              |  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| |
Db      45253 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC 45194

Qy        513 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              ||  |      ||  ||  ||| ||  | | ||| ||
Db      45193 TCTTTCTTTCTCTCTCTCTCTCTCTTTCTTTCTTTCT 45157


RESULT 25
US-09-657-474-7/c
; Sequence 7, Application US/09657474
; Patent No. 6399762
; GENERAL INFORMATION:
;  APPLICANT: Chen, H.
;  APPLICANT:  Freimer, N.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
;  TITLE OF INVENTION:  AND TREATING CHROMOSOME-18p RELATED DISORDERS
;  FILE REFERENCE: 7853-138
;  CURRENT APPLICATION NUMBER: US/09/657,474
;  CURRENT FILING DATE:  2000-09-07
;  PRIOR APPLICATION NUMBER: 09/268,992
;  PRIOR FILING DATE: 1999-03-16
```

```
;  PRIOR APPLICATION NUMBER: 09/236,134
;  PRIOR FILING DATE: 1999-01-22
;  PRIOR APPLICATION NUMBER: 60/106,056
;  PRIOR FILING DATE: 1998-10-28
;  PRIOR APPLICATION NUMBER: 60/088,312
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/078,044
;  PRIOR FILING DATE: 1998-03-16
;  NUMBER OF SEQ ID NOS: 84
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 72604
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: all n positions
;   OTHER INFORMATION: n=a, c, g, or t
US-09-657-474-7

  Query Match             5.1%;  Score 51.4;  DB 4;  Length 72604;
  Best Local Similarity   58.0%;  Pred. No. 0.04;
  Matches   91;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy        393 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 452
              ||  |  | | |||  || ||| ||  | | ||| ||| ||  ||  | | ||| ||| |
Db      45313 TCTTTCTCTTTCTTCTTTTCTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTYCTTYC 45254

Qy        453 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 512
              |  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| |
Db      45253 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC 45194

Qy        513 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              ||  |      ||  ||  ||| ||  | | ||| ||
Db      45193 TCTTTCTTTCTCTCTCTCTCTCTCTTTCTTTCTTTCT 45157


RESULT 26
US-08-903-800A-3
; Sequence 3, Application US/08903800A
; Patent No. 5935789
;  GENERAL INFORMATION:
;    APPLICANT:  RHEE, Sang-Ki
;    APPLICANT:  CHOI, Eui-Sung
;    APPLICANT:  KIM, Chul-Ho
;    APPLICANT:  SOHN, Jung-Hoon
;    APPLICANT:  KANG, Hyun-Ah
;    APPLICANT:  KIM, Hwa-Young
;    TITLE OF INVENTION:  AUTONOMOUSLY REPLICATING SEQUENCES,
;    TITLE OF INVENTION:  GAPDH GENE AND PROMOTER DERIVED FROM
;    TITLE OF INVENTION:  HANSENULA POLYMORPHA, EXPRESSION
;    TITLE OF INVENTION:  VECTORS CONTAINING SAME AND METHOD FOR
;    TITLE OF INVENTION:  THE SELECTION OF TRANSFORMANTS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RHEE, Sang-Ki
;      STREET:  Keukdong Villa Ka-101, Kwangjang-dong,
;      STREET:  Kwangjin-gu
;      CITY:  Seoul
;      STATE:  Seoul
;      COUNTRY:  Republic of Korea
;      ZIP:  143-210
;      ADDRESSEE:  CHOI, Eui-Sung
;      STREET:  Dasol Apt. 102-507, #395-3, Kung-dong,
;      STREET:  Yuseong-gu
;      CITY:  Daejeon
;      STATE:  Daejeon
;      COUNTRY:  Republic of Korea
;      ZIP:  305-335
;      ADDRESSEE:  KIM, Chul-Ho
;      STREET:  Shindonga Apt. 2-1101, Yongjeon-dong, Dong-gu
;      CITY:  Daejeon
;      STATE:  Daejeon
;      COUNTRY:  Republic of Korea
;      ZIP:  300-200
;      ADDRESSEE:  SOHN, Jung-Hoon
;      STREET:  Nuri Apt. 103-506, Wolpyung-dong, Seo-gu
;      CITY:  Daejeon
;      STATE:  Daejeon
;      COUNTRY:  Republic of Korea
;      ZIP:  302-280
;      ADDRESSEE:  KANG, Hyun-Ah
;      STREET:  Kyungseong Keunmaul Apt. 125-1501, Kalma-dong,
;      STREET:  Seo-gu
;      CITY:  Daejeon
;      STATE:  Daejeon
;      COUNTRY:  Republic of Korea
;      ZIP:  302-171
;      ADDRESSEE:  KIM, Hwa-Young
;      STREET:  #3-462, Bukahyun-3-dong, Seodaemun-gu
;      CITY:  Seoul
;      STATE:  Seoul
;      COUNTRY:  Republic of Korea
;      ZIP:  120-190
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5inch, 1.44MB storage
;      COMPUTER:  IBM PC/AT
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/903,800A
;      FILING DATE:  31-JUL-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  KR 97-3173
;      FILING DATE:  31-JAN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:
;      REGISTRATION NUMBER:
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:
;      TELEFAX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  405 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      OTHER INFORMATION:  autonomously replicating sequnce of
;      OTHER INFORMATION:  Hansenula polymorpha DL-1(ATCC 26012)
US-08-903-800A-3

  Query Match             5.0%;  Score 50.4;  DB 2;  Length 405;
  Best Local Similarity   49.6%;  Pred. No. 0.039;
  Matches  129;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        133 GGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGT 192
              ||   ||| || | | ||  | ||||    ||||||  ||| | ||     ||     |
Db        145 GGAAGCGGCAGGAAGCGGTGTAGGGATGCGGCGAGGAAAGCAGAGGCAGCTGGTTTCAGG 204

Qy        193 GTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAG 252
                |||  ||   |         ||   ||  ||  |   ||  | ||||  | | || | |
Db        205 ACGCGGTCTGAGGCCTGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGG 264

Qy        253 GGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGC 312
                ||   |  |||  | | ||  | ||||  | | || | | ||  |  |||  | | ||
Db        265 TGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGC 324

Qy        313 TGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGC 372
              | | ||   |  |||  | | ||  | ||||  | | || | | ||   |  ||   |
Db        325 GGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGGTGGCGGGG 384

Qy        373 TGGCTTGCTAGTAGGGCTGC 392
```

```
                ||||  | | |  |||  ||
Db        385 TGGCGGGGTGGCGGGGTGGC 404


RESULT 27
US-08-623-906A-14/c
; Sequence 14, Application US/08623906A
; Patent No. 5874217
;  GENERAL INFORMATION:
;    APPLICANT:  Stevenson, Tamara
;    APPLICANT:  Dvorak, Jan
;    APPLICANT:  Halverson, Joy
;    TITLE OF INVENTION:  Microsatellite Sequences for Canine
;    TITLE OF INVENTION:  Genotyping
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/623,906A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sherwood, Pamela J.
;      REGISTRATION NUMBER:  36,677
;      REFERENCE/DOCKET NUMBER:  A-62282/BIR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  348 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..21
;      OTHER INFORMATION:  /note= "Nucleotides 1-21 are unique
;      OTHER INFORMATION:  flanking sequence"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  22..156
;      OTHER INFORMATION:  /note= "Nucleotides 22-156 are
;      OTHER INFORMATION:  repeat sequence"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  157..348
;      OTHER INFORMATION:  /note= "Nucleotides 157-348 are
;      OTHER INFORMATION:  unique flanking sequence"
US-08-623-906A-14

  Query Match             5.0%;  Score 49.8;  DB 2;  Length 348;
  Best Local Similarity   57.3%;  Pred. No. 0.051;
  Matches   90;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        400 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 459
              | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||
Db        157 CTTTCTTTCTTCCTTTCTTTCTTTCTTCCTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTT 98

Qy        460 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 519
              | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||   |
Db         97 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTCTTTC 38

Qy        520 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 556
              |  || ||| ||  | || |   || |  || | |
Db         37 TTTCTTTCCTTTCTTTCAGGTCATTTTACAAAGAGCA 1


RESULT 28
US-09-007-005-17/c
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

  Query Match             4.9%;  Score 49.2;  DB 3;  Length 289;
  Best Local Similarity   5.4%;  Pred. No. 0.066;
  Matches   12;  Conservative  109;  Mismatches  101;  Indels    0;  Gaps    0;

Qy        328 CCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGG 387
              :|:|: : : : :| |::: : :::    :     ::    :: :  :: : :   :  :
Db        235 YCYGYCYAYGYCYTYGYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYS 176

Qy        388 GCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 447
                : : : :  :: : : : : ::: : :   : ::: : : : : : :::   ::: : ::
Db        175 YNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYS 116

Qy        448 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 507
              : : : : :  :: : : : : ::: : :   : ::: : : : : : :::   ::: : :
Db        115 YNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYC 56

Qy        508 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              : : :|:|:|  : : :|:|:|: : : :|  : : : : :|
Db         55 YAYTYTYGYTYAYAYTYTYGYTYAYAYAYTYAYGYTYAYAYT 14


RESULT 29
US-09-244-796-17/c
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350007
;  CURRENT APPLICATION NUMBER: US/09/244,796
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
```

```
;' EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  EARLIER APPLICATION NUMBER: 09/007,005
;  EARLIER FILING DATE: 1998-01-14
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

  Query Match             4.9%;  Score 49.2;  DB 3;  Length 289;
  Best Local Similarity   5.4%;  Pred. No. 0.066;
  Matches   12;  Conservative  109;  Mismatches  101;  Indels    0;  Gaps    0;

Qy        328 CCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGG 387
              :|:|: : : : :| |::: : :::    :     ::    :: : :: : :   :  :
Db        235 YCYGYCYAYGYCYTYGYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYS 176

Qy        388 GCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 447
              : : : :  :: : : : : : ::: : :   : ::: : : : : : :::   ::: : ::
Db        175 YNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYS 116

Qy        448 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 507
              : : : : :  :: : : : : : ::: : :   : ::: : : : : : :::   ::: : :
Db        115 YNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYC 56

Qy        508 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              : : :|:|:|  : : :|:|:|: : : :|  : : : : :|
Db         55 YAYTYTYGYTYAYAYTYTYGYTYAYAYAYTYAYGYTYAYAYT 14


RESULT 30
US-08-229-515A-13/c
; Sequence 13, Application US/08229515A
; Patent No. 5518885
;  GENERAL INFORMATION:
;    APPLICANT:  RAZIUDDIN
;    APPLICANT:  SARKAR, FAZLUL H
;    TITLE OF INVENTION:  ERBB2 PROMOTER BINDING PROTEIN IN
;    TITLE OF INVENTION:  NEOPLASTIC DISEASE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG PC
;      STREET:  127 Peachtree Street, Suite 1200
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  usa
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/229,515A
;      FILING DATE:  19 APR 1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PERRYMAN, DAVID G
;      REGISTRATION NUMBER:  33,438
;      REFERENCE/DOCKET NUMBER:  1414.608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  404-688-0770
;      TELEFAX:  404-688-9880
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  650 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-229-515A-13

  Query Match             4.9%;  Score 48.8;  DB 1;  Length 650;
  Best Local Similarity   73.8%;  Pred. No. 0.086;
  Matches   62;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy        399 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 458
              ||||||    |  | ||  | |  |||||||||| ||| || |||||||||||| |||
Db        183 GCCTGCCCTGTGCCCCGACGCCTAGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 124

Qy        459 GCCTGCTTGCTTGCTCGCTGGCCT 482
              || ||||||||||| || |  ||
Db        123 GCTTGCTTGCTTGCTTGCAGTTCT 100


RESULT 31
US-08-645-865-13/c
; Sequence 13, Application US/08645865
; Patent No. 5654406
;  GENERAL INFORMATION:
;    APPLICANT:  RAZIUDDIN
;    APPLICANT:  SARKAR, FAZLUL H
;    TITLE OF INVENTION:  ERBB2 PROMOTER BINDING PROTEIN IN
;    TITLE OF INVENTION:  NEOPLASTIC DISEASE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG PC
;      STREET:  127 Peachtree Street, Suite 1200
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  usa
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/645,865
;      FILING DATE:  14 MAY 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PERRYMAN, DAVID G
;      REGISTRATION NUMBER:  33,438
;      REFERENCE/DOCKET NUMBER:  1414.608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  404-688-0770
;      TELEFAX:  404-688-9880
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  650 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-645-865-13

  Query Match             4.9%;  Score 48.8;  DB 1;  Length 650;
  Best Local Similarity   73.8%;  Pred. No. 0.086;
  Matches   62;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy        399 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 458
              ||||||    |  | ||  | |  |||||||||| ||| || |||||||||||| |||
Db        183 GCCTGCCCTGTGCCCCGACGCCTAGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 124

Qy        459 GCCTGCTTGCTTGCTCGCTGGCCT 482
              || ||||||||||| || |  ||
Db        123 GCTTGCTTGCTTGCTTGCAGTTCT 100
```

```
RESULT 32
US-09-318-443-7/c
; Sequence 7, Application US/09318443
; Patent No. 6197947
; GENERAL INFORMATION:
;  APPLICANT: Hemmati-Brivanlou, Ali
;  APPLICANT: Weinstein, Daniel C.
;  TITLE OF INVENTION: TRANSLATION INITIATION FACTOR 4AIII, AND METHODS OF USE
;  TITLE OF INVENTION: THEREOF
;  FILE REFERENCE: 600-1-211 N
;  CURRENT APPLICATION NUMBER: US/09/318,443
;  CURRENT FILING DATE: 1999-05-25
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 1682
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-318-443-7

  Query Match             4.9%;  Score 48.8;  DB 3;  Length 1682;
  Best Local Similarity   61.7%;  Pred. No. 0.094;
  Matches   95;  Conservative    0;  Mismatches   57;  Indels    2;  Gaps    1;

Qy        392 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 451
              ||||||| | ||| | ||||||| | ||| | ||||||| | ||| |
Db        153 CTCGCTGTGCCGCTGCCGACCTCGCTG--CCGCTGCCGACCTCGCTGTGCCGCTGCCGAC 96

Qy        452 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 511
              ||||||| | ||| | ||||||| | ||| | ||||||| | ||| |
Db         95 CTCGCTGTGCCGCTGCCGACCTCGCTGTGCCGCTGCCGACCTCGCTGTGCCGCTGCCGAC 36

Qy        512 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 545
              ||||||| | ||| | ||||||| | |||
Db         35 CTCGCTGTGCCGCTGCCGACCTCGCTGTGCCGCT 2


RESULT 33
US-09-579-664B-5
; Sequence 5, Application US/09579664B
; Patent No. 6514719
; GENERAL INFORMATION:
;  APPLICANT: Immunex Corporation
;  APPLICANT:  Bird, Timothy A.
;  APPLICANT:  Virca, G. Duke
;  APPLICANT:  Martin, Unja
;  APPLICANT:  Anderson, Dirk M.
;  TITLE OF INVENTION: NOVEL MURINE AND HUMAN KINASES
;  FILE REFERENCE: 2923-A
;  CURRENT APPLICATION NUMBER: US/09/579,664B
;  CURRENT FILING DATE:  2000-05-26
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
;   LENGTH: 3228
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-579-664B-5

  Query Match             4.8%;  Score 47.8;  DB 4;  Length 3228;
  Best Local Similarity   60.3%;  Pred. No. 0.16;
  Matches   79;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy        404 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 463
              ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | |
Db       3096 CTTTCTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 3155

Qy        464 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 523
              ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | |
Db       3156 CTTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 3215

Qy        524 CTTGCTTGCTC 534
              ||| ||  ||
Db       3216 CTTCTTTCTTC 3226


RESULT 34
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             4.8%;  Score 47.8;  DB 1;  Length 7218;
  Best Local Similarity   8.8%;  Pred. No. 0.17;
  Matches   19;  Conservative  123;  Mismatches   75;  Indels    0;  Gaps    0;

Qy        333 TTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGC 392
              || |  ||| | ||||  |  |||||  |    |  :: ::  ::: ::  :    :: :
Db       1031 TTCCGAGCTTGGCTGCAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYY 1090

Qy        393 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 452
              :: ::  ::: ::: ::: ::: ::  ::: ::: ::: ::: ::  ::: ::: ::: :
Db       1091 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1150

Qy        453 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 512
              :: ::  ::: ::: ::: ::: ::  ::: ::: ::: ::: ::  ::: ::: ::: :
Db       1151 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1210

Qy        513 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              :: ::  ::: ::: ::: ::: ::  ::: ::: ::
Db       1211 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1247
```

RESULT 35
US-09-252-991A-10276/c
; Sequence 10276, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10276
; LENGTH: 852
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10276

Query Match 4.8%; Score 47.6; DB 4; Length 852;
Best Local Similarity 49.6%; Pred. No. 0.15;
Matches 122; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Qy 1 CGCCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTG 60
Db 503 CGCCTCGCCGAACGCCTGACCACCCTGCGCCTGCTCAGCCGCCCGGCGGCGGAGATCCGC 444

Qy 61 GTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGG 120
Db 443 GCCCAGGCCGAGCGTCTGGCGCCGGCGCTGGGCGAGGCGCTCGGCGAGGGCTGGGAGGTC 384

Qy 121 GGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGAC 180
Db 383 GCCGTGGTCGACGCCCCTGGGCATGATCGGCAGCGGCGCGCAGCCGGTGGCGCGGCTGCC 324

Qy 181 AGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGC 240
Db 323 CAGCGCCGCGCTGTGCCTGCGTCCGCGGCAGCCCAGGCGCCTGCGCGGACGCGCCCTGCG 264

Qy 241 CTGCTT 246
Db 263 CAACCT 258

RESULT 36
US-09-252-991A-10486/c
; Sequence 10486, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10486
; LENGTH: 2778
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10486

Query Match 4.8%; Score 47.6; DB 4; Length 2778;
Best Local Similarity 49.6%; Pred. No. 0.17;
Matches 122; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Qy 1 CGCCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTG 60
Db 1731 CGCCTCGCCGAACGCCTGACCACCCTGCGCCTGCTCAGCCGCCCGGCGGCGGAGATCCGC 1672

Qy 61 GTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGG 120
Db 1671 GCCCAGGCCGAGCGTCTGGCGCCGGCGCTGGGCGAGGCGCTCGGCGAGGGCTGGGAGGTC 1612

Qy 121 GGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGAC 180
Db 1611 GCCGTGGTCGACGCCCCTGGGCATGATCGGCAGCGGCGCGCAGCCGGTGGCGCGGCTGCC 1552

Qy 181 AGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGC 240
Db 1551 CAGCGCCGCGCTGTGCCTGCGTCCGCGGCAGCCCAGGCGCCTGCGCGGACGCGCCCTGCG 1492

Qy 241 CTGCTT 246
Db 1491 CAACCT 1486

RESULT 37
US-09-252-991A-10685
; Sequence 10685, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10685
; LENGTH: 3339
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10685

Query Match 4.8%; Score 47.6; DB 4; Length 3339;
Best Local Similarity 49.6%; Pred. No. 0.17;
Matches 122; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Qy 1 CGCCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTG 60
Db 1033 CGCCTCGCCGAACGCCTGACCACCCTGCGCCTGCTCAGCCGCCCGGCGGCGGAGATCCGC 1092

Qy 61 GTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGG 120
Db 1093 GCCCAGGCCGAGCGTCTGGCGCCGGCGCTGGGCGAGGCGCTCGGCGAGGGCTGGGAGGTC 1152

Qy 121 GGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGAC 180
Db 1153 GCCGTGGTCGACGCCCCTGGGCATGATCGGCAGCGGCGCGCAGCCGGTGGCGCGGCTGCC 1212

Qy 181 AGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGC 240
Db 1213 CAGCGCCGCGCTGTGCCTGCGTCCGCGGCAGCCCAGGCGCCTGCGCGGACGCGCCCTGCG 1272

Qy 241 CTGCTT 246
Db 1273 CAACCT 1278

RESULT 38
US-09-268-992-7
; Sequence 7, Application US/09268992
; Patent No. 6342351

```
; GENERAL INFORMATION:
;  APPLICANT: Chen, H.
;  APPLICANT: Freimer, N.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
;  TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
;  FILE REFERENCE: 7853-138
;  CURRENT APPLICATION NUMBER: US/09/268,992
;  CURRENT FILING DATE: 1999-03-16
;  EARLIER APPLICATION NUMBER: 09/236,134
;  EARLIER FILING DATE: 1999-01-22
;  EARLIER APPLICATION NUMBER: 60/106,056
;  EARLIER FILING DATE: 1998-10-28
;  EARLIER APPLICATION NUMBER: 60/088,312
;  EARLIER FILING DATE: 1998-06-05
;  EARLIER APPLICATION NUMBER: 60/078,044
;  EARLIER FILING DATE: 1998-03-16
;  NUMBER OF SEQ ID NOS: 84
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 72604
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: all n positions
;   OTHER INFORMATION: n=a, c, g, or t
US-09-268-992-7

  Query Match             4.7%;  Score 47.4;  DB 4;  Length 72604;
  Best Local Similarity   61.7%;  Pred. No. 0.25;
  Matches   92;  Conservative    0;  Mismatches   56;  Indels    1;  Gaps    1;

Qy        401 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 460
              || ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db       7174 CTCCTTCCTTCCTTCCTTCCTTTCTTTCTTTCT-CCTTCCTTCCTTCCTTCCTTCC 7232

Qy        461 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 520
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db       7233 TTCCTTCCTTCCTTCCTTCCTTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC 7292

Qy        521 CTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              | ||| ||| ||  ||  ||| ||| ||
Db       7293 TTTCTTTCTTTCTTTCTTTCCTTCTTTCT 7321


RESULT 39
US-09-657-474-7
; Sequence 7, Application US/09657474
; Patent No. 6399762
; GENERAL INFORMATION:
;  APPLICANT: Chen, H.
;  APPLICANT:  Freimer, N.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
;  TITLE OF INVENTION:  AND TREATING CHROMOSOME-18p RELATED DISORDERS
;  FILE REFERENCE: 7853-138
;  CURRENT APPLICATION NUMBER: US/09/657,474
;  CURRENT FILING DATE:  2000-09-07
;  PRIOR APPLICATION NUMBER: 09/268,992
;  PRIOR FILING DATE: 1999-03-16
;  PRIOR APPLICATION NUMBER: 09/236,134
;  PRIOR FILING DATE: 1999-01-22
;  PRIOR APPLICATION NUMBER: 60/106,056
;  PRIOR FILING DATE: 1998-10-28
;  PRIOR APPLICATION NUMBER: 60/088,312
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/078,044
;  PRIOR FILING DATE: 1998-03-16
;  NUMBER OF SEQ ID NOS: 84
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 72604
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: all n positions
;   OTHER INFORMATION: n=a, c, g, or t
US-09-657-474-7

  Query Match             4.7%;  Score 47.4;  DB 4;  Length 72604;
  Best Local Similarity   61.7%;  Pred. No. 0.25;
  Matches   92;  Conservative    0;  Mismatches   56;  Indels    1;  Gaps    1;

Qy        401 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 460
              || ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db       7174 CTCCTTCCTTCCTTCCTTCCTTCCTTTCTTTCTTTCT-CCTTCCTTCCTTCCTTCCTTCC 7232

Qy        461 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 520
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db       7233 TTCCTTCCTTCCTTCCTTCCTTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC 7292

Qy        521 CTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              | ||| ||| ||  ||  ||| ||| ||
Db       7293 TTTCTTTCTTTCTTTCTTTCCTTCTTTCT 7321


RESULT 40
US-09-165-264-7
; Sequence 7, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
;  APPLICANT:  Vinayagamoorthy, Thuraiayah
;  TITLE OF INVENTION:  Multi-Loci Genomic Analysis
;  FILE REFERENCE:  44747
;  CURRENT APPLICATION NUMBER: US/09/165,264
;  CURRENT FILING DATE:  1998-10-01
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 7
;   LENGTH: 320
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-7

  Query Match             4.7%;  Score 47.2;  DB 3;  Length 320;
  Best Local Similarity   48.2%;  Pred. No. 0.17;
  Matches  133;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy         57 GGTGGTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGC 116
              || ||      | |  |  ||| ||||||  ||||   ||||  |    || ||||   ||
Db         38 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 97

Qy        117 GCGGGGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGG 176
              | |||| | |   | |||    ||| |   |||||  ||||     | | |||| | |||
Db         98 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 157

Qy        177 GGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGC 236
              ||   ||||  || | | | | |   | | |     |||   ||  |   |   |   |
Db        158 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 217

Qy        237 TGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 296
              ||   |   | | ||||   | ||   |   |   |  ||   |   | | ||||
Db        218 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 277

Qy        297 CAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGC 332
                |  ||   |   |  | |||||   ||| | |||
Db        278 GGGGGGGGGGGGGGGGGGGGGGAGGCTAGCCATGC 313


RESULT 41
US-09-165-264-13
; Sequence 13, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
```

```
;  APPLICANT:  Vinayagamoorthy, Thuraiayah
;  TITLE OF INVENTION:  Multi-Loci Genomic Analysis
;  FILE REFERENCE:  44747
;  CURRENT APPLICATION NUMBER: US/09/165,264
;  CURRENT FILING DATE:  1998-10-01
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 13
;   LENGTH: 320
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-13

  Query Match             4.7%;  Score 47.2;  DB 3;  Length 320;
  Best Local Similarity   47.6%;  Pred. No. 0.17;
  Matches  139;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy         57 GGTGGTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGC 116
              || ||     | |  |  ||| ||||||  ||||   |||| |    || ||||   ||
Db         27 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 86

Qy        117 GCGGGGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGG 176
              | |||| | |   | |||    ||| |   |||||   |||     | | |||| | |||
Db         87 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 146

Qy        177 GGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGC 236
              ||   ||||  || | | | | |   | | |     |||   ||  |   |   |   |
Db        147 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 206

Qy        237 TGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 296
               ||   |   | | ||||   |  ||   |   |   |  ||   |   |  | ||||
Db        207 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 266

Qy        297 CAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGC 348
                 |  ||   |   | |||||   |  ||    | ||  ||  ||| |||
Db        267 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGACGTGGACGAGGGCATGC 318


RESULT 42
US-08-586-165-1/c
; Sequence 1, Application US/08586165
; Patent No. 6054298
;  GENERAL INFORMATION:
;    APPLICANT:  Laufer, Edward M.
;    APPLICANT:  Orozco, Olivia E.
;    APPLICANT:  Tabin, Clifford J.
;    TITLE OF INVENTION:  Fringe Proteins and Pattern Formation
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/586,165
;      FILING DATE:  16-JAN-1996
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  HU95-05
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 861-6240
;      TELEFAX:  (617) 861-9540
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1120 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1..240, 244..474, 478..531, 538..579, 583
;      LOCATION:  ..678, 682..687, 691..807, 811..843, 850..870,
;      LOCATION:  874..990, 994..1056, 1060..1083, 1087..1104, 1108
;      LOCATION:  ..1119)
US-08-586-165-1

  Query Match             4.7%;  Score 46.6;  DB 3;  Length 1120;
  Best Local Similarity   56.1%;  Pred. No. 0.25;
  Matches   88;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy          3 CCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGT 62
              || ||  |  |  ||  || ||||||  |||||  | |   | |||||  |  |
Db        221 CCCCCGCGGCTCCCGAGCCGCCGCCGCGGTCCCCGGGCACGTCGCTCCCCGCGGGCCGCG 162

Qy         63 CTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGG 122
              |  | |||  | |||||||    | ||||||| | ||  |||  ||   ||| ||| ||
Db        161 CCTCCCGCTTCGGGGAGGGCCTGCTGGGCCCGGCGGGCGGCGGGCGGCGGCGGGGCGCGG 102

Qy        123 CGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAA 159
              || |     |  |||.||| || | |  |   | |||
Db        101 CGGGGGGCTGTCCCCGGGGCGAGCAGCAGCAGGAGAA 65


RESULT 43
US-08-765-875-1
; Sequence 1, Application US/08765875
; Patent No. 5914234
;  GENERAL INFORMATION:
;    APPLICANT:  LEE, SE-JIN
;    APPLICANT:  MCPHERRON, ALEXANDRA C.
;    TITLE OF INVENTION:  GROWTH DIFFERENTIATION FACTOR-11
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SPENSLEY HORN JUBAS & LUBITZ
;      STREET:  1880 CENTURY PARK EAST, FIFTH FLOOR
;      CITY:  LOS ANGELES
;      STATE:  CALIFORNIA
;      COUNTRY:  US
;      ZIP:  90067
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/765,875
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/706,958
;      FILING DATE:
;      APPLICATION NUMBER:  US/08/272,763
;      FILING DATE:  08-JUL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  TUMARKIN PH.D., LISA A.
;      REGISTRATION NUMBER:  P-38,347
;      REFERENCE/DOCKET NUMBER:  PD3641
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619/455-5100
;      TELEFAX:  619/455-5110
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1393 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  HUMAN GDF-11
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  54..1274
US-08-765-875-1

  Query Match             4.7%;  Score 46.6;  DB 2;  Length 1393;
  Best Local Similarity   55.2%;  Pred. No. 0.25;
  Matches   91;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy         14 GGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCTCCGCGCCGC 73
              ||||  ||  | ||    ||| | | |||   ||  || | | |||| || | || ||||
Db         13 GGCGTCCCCGCCCCCCAGTCCTCCCTCCCCTCCCCTCCAGCATGGTGCTCGCGGCCCCGC 72

Qy         74 CGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGACCGAGG 133
              |     |||       || |   || |     || | | ||   | |||  ||| ||||||
Db         73 TGCTGCTGGGCTTCCTGCTCCTCGCCCTGGAGCTGCGGCCCCGGGGGGAGGCGGCCGAGG 132

Qy        134 GCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGG 178
              ||||||    ||| ||    ||     |||  || ||||| ||
Db        133 GCCCCGCGGCGGCGGCGGCGGCGGCGGCGGCAGCGGCGG 177


RESULT 44
US-08-795-671-1
; Sequence 1, Application US/08795671
; Patent No. 6008434
;  GENERAL INFORMATION:
;    APPLICANT:  Se-Jin Lee and Alexandra McPherron
;    TITLE OF INVENTION:  GROWTH DIFFERENTIATION FACTOR-11
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/795,671
;      FILING DATE:  February 6, 1997
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  HAILE, PH.D., LISA A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  07265/106001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619/678-5070
;      TELEFAX:  619/678-5099
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1393 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  HUMAN GDF-11
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  54..1274
US-08-795-671-1

  Query Match             4.7%;  Score 46.6;  DB 3;  Length 1393;
  Best Local Similarity   55.2%;  Pred. No. 0.25;
  Matches   91;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy         14 GGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCTCCGCGCCGC 73
              ||||  ||  | ||    ||| | | |||   ||  || | | |||| || | || ||||
Db         13 GGCGTCCCCGCCCCCCAGTCCTCCCTCCCCTCCCCTCCAGCATGGTGCTCGCGGCCCCGC 72

Qy         74 CGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGACCGAGG 133
              |     |||       || |   || |     || | | ||   | |||  ||| ||||||
Db         73 TGCTGCTGGGCTTCCTGCTCCTCGCCCTGGAGCTGCGGCCCCGGGGGGAGGCGGCCGAGG 132

Qy        134 GCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGG 178
              ||||||    ||| ||    ||     |||  || ||||| ||
Db        133 GCCCCGCGGCGGCGGCGGCGGCGGCGGCGGCAGCGGCGG 177


RESULT 45
US-09-454-540-1
; Sequence 1, Application US/09454540
; Patent No. 6517835
;  GENERAL INFORMATION:
;    APPLICANT:  Se-Jin Lee and Alexandra McPherron
;    TITLE OF INVENTION:  GROWTH DIFFERENTIATION FACTOR-11
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/454,540
;      FILING DATE:  06-DEC-1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/795,671
;      FILING DATE:  February 6, 1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  HAILE, PH.D., LISA A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  07265/106001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619/678-5070
;      TELEFAX:  619/678-5099
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1393 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  HUMAN GDF-11
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  54..1274
US-09-454-540-1

  Query Match             4.7%;  Score 46.6;  DB 4;  Length 1393;
  Best Local Similarity   55.2%;  Pred. No. 0.25;
  Matches   91;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy         14 GGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCTCCGCGCCGC 73
              ||||  ||  | ||    ||| | | |||   ||  || | | |||| || | || ||||
Db         13 GGCGTCCCCGCCCCCCAGTCCTCCCTCCCCTCCCCTCCAGCATGGTGCTCGCGGCCCCGC 72

Qy         74 CGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGACCGAGG 133
```

```
Db         73 TGCTGCTGGGCTTCCTGCTCCTCGCCCTGGAGCTGCGGCCCCGGGGGGAGGCGGCCGAGG 132

Qy        134 GCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGG 178
              |||||| |    ||| ||    ||     |||  || ||||| ||
Db        133 GCCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCAGCGGCGG 177


RESULT 46
US-07-753-520B-2/c
; Sequence 2, Application US/07753520B
; Patent No. 5352595
;  GENERAL INFORMATION:
;    APPLICANT:  Tapscott,J.;Weintraub,H.M.;Palmer,T.D.
;    TITLE OF INVENTION:  "MyoD REGULATORY REGION"
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;      STREET:  2800 Pacific First Center, 1420 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-5.25 inch, 1.2Mb storage
;      COMPUTER:  IBM PC/386 Compatible
;      OPERATING SYSTEM:  MS-DOS 4.01
;      SOFTWARE:  Word for Windows-t
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/753,520B
;      FILING DATE:  19910903
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  none
;      FILING DATE:  none
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sundsmo,John,S.
;      REGISTRATION NUMBER:  34,446
;      REFERENCE/DOCKET NUMBER:  FHCR-1-5789
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  1-206-682-8100; 1-206-224-0727 (direct)
;      TELEFAX:  1-206-224-0779
;      TELEX:  4938023
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1978 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;      DESCRIPTION:  KX fragment MyoD Enhancer; distal regulatory region myoD gene; Figu
US-07-753-520B-2

  Query Match             4.7%;  Score 46.6;  DB 1;  Length 1978;
  Best Local Similarity   53.6%;  Pred. No. 0.26;
  Matches   97;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy        378 TGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 437
              ||| | |   | | ||  ||  | | ||| ||| ||  ||  ||  | | ||| ||| ||  ||
Db       1382 TGCAAATCTAGTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 1323

Qy        438 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 497
                | | ||| ||| ||  ||  | | ||| ||  ||| ||  |   ||  ||  ||| ||
Db       1322 TTCTTCCTTCCTTCCTTCCTTCCTTCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1263

Qy        498 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGAC 557
                  ||  ||  ||| ||  |   ||  ||  ||| ||   ||      | |||   ||
Db       1262 CCTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTATCTCAAACCCGAGCAAAC 1203

Qy        558 A 558
              |
Db       1202 A 1202


RESULT 47
US-08-586-165-8/c
; Sequence 8, Application US/08586165
; Patent No. 6054298
;  GENERAL INFORMATION:
;    APPLICANT:  Laufer, Edward M.
;    APPLICANT:  Orozco, Olivia E.
;    APPLICANT:  Tabin, Clifford J.
;    TITLE OF INVENTION:  Fringe Proteins and Pattern Formation
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/586,165
;      FILING DATE:  16-JAN-1996
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  HU95-05
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 861-6240
;      TELEFAX:  (617) 861-9540
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3937 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-586-165-8

  Query Match             4.7%;  Score 46.6;  DB 3;  Length 3937;
  Best Local Similarity   56.1%;  Pred. No. 0.28;
  Matches   88;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy          3 CCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGT 62
              || ||  |  |  ||  || ||||||   |||||   | |    | |||||   |  |
Db        403 CCCCCGCGGCTCCCGAGCCGCCGCCGCGGTCCCCGGGCACGTCGCTCCCCGCGGGCCGCG 344

Qy         63 CTCCGCGCCGCCGGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGG 122
              |  | |||  | |||||||    | |||||||| | ||  |||  ||   ||| ||| ||
Db        343 CCTCCCGCTTCGGGGAGGGCCTGCTGGGCCCGGCGGGCGGCGGGCGGCGGCGGGGCGCGG 284

Qy        123 CGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAA 159
              || |     |  ||| ||| || | |  |   | |||
Db        283 CGGGGGGCTGTCCCCGGGGCGAGCAGCAGCAGGAGAA 247


RESULT 48
US-07-753-520B-3/c
; Sequence 3, Application US/07753520B
; Patent No. 5352595
;  GENERAL INFORMATION:
;    APPLICANT:  Tapscott,J.;Weintraub,H.M.;Palmer,T.D.
;    TITLE OF INVENTION:  "MyoD REGULATORY REGION"
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;      STREET:  2800 Pacific First Center, 1420 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
```

```
;       COUNTRY:  USA
;       ZIP:  98101-2347
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette-5.25 inch, 1.2Mb storage
;       COMPUTER:  IBM PC/386 Compatible
;       OPERATING SYSTEM:  MS-DOS 4.01
;       SOFTWARE:  Word for Windows-t
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/753,520B
;       FILING DATE:  19910903
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  none
;       FILING DATE:  none
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sundsmo,John,S.
;       REGISTRATION NUMBER:  34,446
;       REFERENCE/DOCKET NUMBER:  FHCR-1-5789
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  1-206-682-8100; 1-206-224-0727 (direct)
;       TELEFAX:  1-206-224-0779
;       TELEX:  4938023
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  9115 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Other;plasmid DNA
;       DESCRIPTION:  pLHDMDN-53: 5'LTR (position 1-1159); Y+ (position
;       DESCRIPTION:  1159-1640); HisD (position 1641-3007); Myo-D 531.4 ApaI fragment
;       DESCRIPTION:  (position 3008-5248); driving neo (position 5249-6117); with a
;       DESCRIPTION:  3'LTR (position 6118-6823) coupled to a pBR322 plasmid (position
;       DESCRIPTION:  6824-9115); Figures 7A-7D.
US-07-753-520B-3

  Query Match             4.7%;  Score 46.6;  DB 1;  Length 9115;
  Best Local Similarity   53.6%;  Pred. No. 0.3;
  Matches   97;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy        378 TGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 437
              ||| | |   | | ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||
Db       4389 TGCAAATCTAGTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTCT 4330

Qy        438 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 497
              | | ||| ||| ||  ||  | | ||| ||  ||| ||  |   ||  ||  ||| ||
Db       4329 TTCTTCCTTCCTTCCTTCCTTCCTTCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 4270

Qy        498 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGAC 557
                  ||  ||  ||| ||  |   ||  ||  ||| ||   ||     | |||   ||
Db       4269 CCTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTATCTCAAACCCGAGCAAAC 4210

Qy        558 A 558
              |
Db       4209 A 4209


RESULT 49
US-09-128-155-16/c
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
;  APPLICANT: Pan, Yang
;  TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
;  TITLE OF INVENTION: AND USES THEREOF
;  FILE REFERENCE: 09404/052001
;  CURRENT APPLICATION NUMBER: US/09/128,155
;  CURRENT FILING DATE: 1998-08-03
;  EARLIER APPLICATION NUMBER: US 60/091,650
;' EARLIER FILING DATE: 1998-07-02
;  EARLIER APPLICATION NUMBER: US 60/054,646
;  EARLIER FILING DATE: 1997-08-04
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 16
;   LENGTH: 152331
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(152331)
;   OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

  Query Match             4.7%;  Score 46.6;  DB 3;  Length 152331;
  Best Local Similarity   49.8%;  Pred. No. 0.39;
  Matches  144;  Conservative    0;  Mismatches  144;  Indels    1;  Gaps    1;

Qy         60 GGTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCG 119
              || | |||||||| |||| ||||||||||||   |||| ||   ||||||||   | | |
Db      22217 GGGCGCCGCGCCGGCGGGGGGGGGGCCCGGG-GGGGGGGCGGGGGCGGGGCGGCGGGGG 22159

Qy        120 GGGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGA 179
              |||   |       ||  | ||| |  ||||||   ||         | |||| | |||||
Db      22158 GGGTCGGGGGTGCGGGGCGGGGCGGGCGGGGGGGGGGCGTCGGCTGCGGGGGGGGGGGGG 22099

Qy        180 CAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGG 239
              | ||||     | ||| | | |          ||||||  ||  |   |   |   |  ||
Db      22098 CGGGGGCGTCGGGGTGGGGGGTGCGTGTGGGCCGGGTGTGGGGGGTGGGGGGGGGGGGGG 22039

Qy        240 CCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAG 299
                ||  |  |  |||   || ||       |   |  || | |   || | ||||   |
Db      22038 GGGGCGGGGGGGCGGGGGGGTGGGGGGCGCGGGGGGGGGGGCGGGCGCGGGGGGGGGG 21979

Qy        300 TAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGC 348
                ||   |   |  |  | |   |  || | |   |  ||  ||   ||
Db      21978 GGGGGGGGTGGGGGGCCGCGGGGGGCGGGCCGGGGGGGCGGGGGGGGC 21930


RESULT 50
US-08-332-766A-37/c
; Sequence 37, Application US/08332766A
; Patent No. 5843647
;  GENERAL INFORMATION:
;    APPLICANT:  JEFFREYS, Alec J.
;    APPLICANT:  ARMOUR, John
;    TITLE OF INVENTION:  SIMPLE TANDEM REPEATS
;    NUMBER OF SEQUENCES:  125
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D. C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/332,766A
;      FILING DATE:  01-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9326052.9
;      FILING DATE:  21-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BIRD, Donald J.
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  217211/M94/0434/GB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  37:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  124 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear.
;    MOLECULE TYPE:  DNA (genomic)
US-08-332-766A-37

  Query Match             4.6%;  Score 46.4;  DB 2;  Length 124;
  Best Local Similarity   72.0%;  Pred. No. 0.22;
  Matches   59;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy        383 GTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 442
              ||   || |||  || ||||||| |||||||| ||| || ||||||||||||| ||| || |
Db         82 GTGACGCCGCTNNCTTGCCTGCCTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTT 23

Qy        443 GCTTGCTTGCTCGCTGGCCTGC 464
                ||| |||  ||  ||  | | |
Db         22 TCTTTCTTTCTTTCTTTCTTTC 1


RESULT 51
US-08-775-009-36
; Sequence 36, Application US/08775009
; Patent No. 5935783
;  GENERAL INFORMATION:
;    APPLICANT:  Gong, Weilong
;    APPLICANT:  Emanuel, Beverly S.
;    APPLICANT:  Budarf, Marcia L.
;    APPLICANT:  Roe, Bruce
;    TITLE OF INVENTION:  No. 5935783el Genes Mapping in the Digeorge and
;    TITLE OF INVENTION:  Veliocardiofacial Syndrome Minimal Critical Region
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz &
;      ADDRESSEE:  No. 5935783ris, LLP
;      STREET:  One Liberty Place - 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  U.S.A.
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/775,009
;      FILING DATE:  27-DEC-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trujillo, Doreen Yatko
;      REGISTRATION NUMBER:  35,719
;      REFERENCE/DOCKET NUMBER:  CH-0681
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3507 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  163..423
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  532..783
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  1351..1458
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(163..423, 532..783, 1351..1458)
US-08-775-009-36

  Query Match             4.6%;  Score 46.4;  DB 2;  Length 3507;
  Best Local Similarity   53.6%;  Pred. No. 0.3;
  Matches  119;  Conservative    0;  Mismatches  101;  Indels    2;  Gaps    1;

Qy          1 CGCCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCC--CACGCCGCCCCGAAAGG 58
              | | ||||||  | | | || |||| | || ||  || ||    | ||||   |  ||
Db        264 CCCGGCCCGGGCCGCCTGCCCACCGCAGCCCGCCGGTCGCCAGAGCCCCGCGAAGCCAGA 323

Qy         59 TGGTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGC 118
               |  | | | ||  ||| |   | | || | ||| |  ||  ||     || ||| ||||
Db        324 GGAGCCCGGGGCGCCCGAGGCTGCGCCCTGCGCCTGCTGCTGCTGCTGCGGCCCCCGCGC 383

Qy        119 GGGGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGG 178
              || || |  | |    ||   || |  ||  ||   |||  |  | | | |||||||||
Db        384 GGCGCCCTGCGGGCCCCCAGAGGCGGCCGCCGGGCTGGGTGAGTGGGCGCGGAGCGGGGC 443

Qy        179 ACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGG 220
                | |||  |||||    |  |   | |   ||  |||| | |
Db        444 GCGGGGCCCGGCGGAGCCCGGGGCGCGGCGCAGTGGGTGCCG 485


RESULT 52
US-09-165-264-11
; Sequence 11, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
;  APPLICANT:  Vinayagamoorthy, Thuraiayah
;  TITLE OF INVENTION:  Multi-Loci Genomic Analysis
;  FILE REFERENCE:  44747
;  CURRENT APPLICATION NUMBER: US/09/165,264
;  CURRENT FILING DATE:  1998-10-01
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 11
;   LENGTH: 320
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-11

  Query Match             4.6%;  Score 46.2;  DB 3;  Length 320;
  Best Local Similarity   46.7%;  Pred. No. 0.27;
  Matches  147;  Conservative    0;  Mismatches  168;  Indels    0;  Gaps    0;

Qy         75 GGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGACCGAGGG 134
              ||| ||||||  ||||   |||| |    || ||||   || | |||| | |   | |||
Db          5 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 64

Qy        135 CCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGTGT 194
                  ||| |   |||||   |||     | | |||| | |||||   |||| || | |
Db         65 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 124

Qy        195 GCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGG 254
              | | |   | | |     |||   ||  |   |   |   |  ||   |   |  | |||
Db        125 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 184

Qy        255 GACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTG 314
              |   |  ||   |   |  | ||   |   |  | ||||   |  ||   |   | |
Db        185 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 244

Qy        315 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTG 374
               ||||   |  ||   |   |   |  ||   |   |  ||||   |  |||  || |
Db        245 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGCAG 304

Qy        375 GCTTGCTAGTAGGGC 389
                ||   || | | |
Db        305 ATTTTGAAGAAAGAC 319
```

```
RESULT 53
US-09-252-991A-16386/c
; Sequence 16386, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 16386
;   LENGTH: 951
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16386

  Query Match             4.6%;  Score 46.2;  DB 4;  Length 951;
  Best Local Similarity   49.2%;  Pred. No. 0.3;
  Matches  212;  Conservative    0;  Mismatches  208;  Indels   11;  Gaps    3;

Qy        103 CGGCGGGGCCCGGCGCGGGGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAAC 162
Db        941 CTGCTCGGCGTCGCCGGAGCGTTCGCTGCCGGAGCGCTGGCAGCCGGTGTGGCGGCGGGG 882

Qy        163 GCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCT 222
Db        881 GCAGCGGGCGCGCTCGAAGACGCGGCGGGAGCGGCCGGTTTCGCCGCGGCGGGCTTGGCT 822

Qy        223 GCCTGCCTGCTCGCTG--GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGC 280
Db        821 GCCGGCTTCTTCGCCGCAGGCTTTTCGCAGCCGGCTTGGCGGCAGGTTTCGCCGCGGGC 762

Qy        281 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGG---GACAGTAGGCCTGCTTGCT 337
Db        761 TTGGCGGCGGTCGCAGCGGCCGGCTTGGCTGCGGCAGGTTTCGCAGCAGGCTTCGCCGCG 702

Qy        338 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCT 397
Db        701 GCCTTGGCTGCCGGCTTGGTCGCCGGTTTCGCTGCGGCCTTCGCAGTCGGTTTGGCGGCG 642

Qy        398 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 457
Db        641 GGCTTGGCTGCCGGCTTGGCGGCCGCGGTTTTCGCCG------CGGGTTTTTTCGCCGCG 588

Qy        458 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 517
Db        587 GGTTTCGCCGCAGGCTTGGCGGCGGCCTTGGCTGCCGGTTTAGCTGCCGGCTTGGCCGCT 528

Qy        518 GGCCTGCTTGC 528
Db        527 GCGGTTTTCGC 517


RESULT 54
US-09-252-991A-16212
; Sequence 16212, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 16212
;   LENGTH: 1092
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16212

  Query Match             4.6%;  Score 46.2;  DB 4;  Length 1092;
  Best Local Similarity   49.2%;  Pred. No. 0.3;
  Matches  212;  Conservative    0;  Mismatches  208;  Indels   11;  Gaps    3;

Qy        103 CGGCGGGGCCCGGCGCGGGGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAAC 162
Db        134 CTGCTCGGCGTCGCCGGAGCGTTCGCTGCCGGAGCGCTGGCAGCCGGTGTGGCGGCGGGG 193

Qy        163 GCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCT 222
Db        194 GCAGCGGGCGCGCTCGAAGACGCGGCGGGAGCGGCCGGTTTCGCCGCGGCGGGCTTGGCT 253

Qy        223 GCCTGCCTGCTCGCTG--GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGC 280
Db        254 GCCGGCTTCTTCGCCGCAGGCTTTTTCGCAGCCGGCTTGGCGGCAGGTTTCGCCGCGGGC 313

Qy        281 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGG---GACAGTAGGCCTGCTTGCT 337
Db        314 TTGGCGGCGGTCGCAGCGGCCGGCTTGGCTGCGGCAGGTTTCGCAGCAGGCTTCGCCGCG 373

Qy        338 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCT 397
Db        374 GCCTTGGCTGCCGGCTTGGTCGCCGGTTTCGCTGCGGCCTTCGCAGTCGGTTTGGCGGCG 433

Qy        398 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 457
Db        434 GGCTTGGCTGCCGGCTTGGCGGCCGCGGTTTTCGCCG------CGGGTTTTTTCGCCGCG 487

Qy        458 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 517
Db        488 GGTTTCGCCGCAGGCTTGGCGGCGGCCTTGGCTGCCGGTTTAGCTGCCGGCTTGGCCGCT 547

Qy        518 GGCCTGCTTGC 528
Db        548 GCGGTTTTCGC 558


RESULT 55
US-09-252-991A-16493/c
; Sequence 16493, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 16493
;   LENGTH: 1311
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16493

  Query Match             4.6%;  Score 46.2;  DB 4;  Length 1311;
  Best Local Similarity   49.2%;  Pred. No. 0.31;
  Matches  212;  Conservative    0;  Mismatches  208;  Indels   11;  Gaps    3;
```

```
Qy        103 CGGCGGGGCCCGGCGCGGGGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAAC 162
              | ||  |||   ||  | |      | |  ||  |   |  | | || |    ||
Db        517 CTGCTCGGCGTCGCCGGAGCGTTCGCTGCCGGAGCGCTGGCAGCCGGTGTGGCGGCGGGG 458

Qy        163 GCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCT 222
              ||   ||| |||   ||     |    | | | |   | ||  |   |  ||||   ||||
Db        457 GCAGCGGGCGCGCTCGAAGACGCGGCGGGAGCGGCCGGTTTCGCCGCGGCGGGCTTGGCT 398

Qy        223 GCCTGCCTGCTCGCTG--GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGC 280
              ||| || |  |||| |  | ||  ||       ||   |  ||| | ||   ||| |||
Db        397 GCCGGCTTCTTCGCCGCAGGCTTTTTCGCAGCCGGCTTGGCGGCAGGTTTCGCCGCGGGC 338

Qy        281 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGG---GACAGTAGGCCTGCTTGCT 337
               ||   || |  |   | |  ||| ||  ||| |  ||     ||| |||| |    ||
Db        337 TTGGCGGCGGTCGCAGCGGCCGGCTTGGCTGCGGCAGGTTTCGCAGCAGGCTTCGCCGCG 278

Qy        338 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCT 397
              || | |||  |||  |  | | ||   |||| | |||   ||| ||  ||   ||
Db        277 GCCTTGGCTGCCGGCTTGGTCGCCGGTTTCGCTGCGGCCTTCGCAGTCGGTTTGGCGGCG 218

Qy        398 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 457
              ||| ||  |||  ||| |  ||||    |  | ||        | | ||  ||   |||
Db        217 GGCTTGGCTGCCGGCTTGGCGGCCGCGGTTTTCGCCG------CGGGTTTTTCGCCGCG 164

Qy        458 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 517
              ||  |    ||  ||| |  |||   |||| |||  | |   |||| |||||  ||||
Db        163 GGTTTCGCCGCAGGCTTGGCGGCGGCCTTGGCTGCCGGTTTAGCTGCCGGCTTGGCCGCT 104

Qy        518 GGCCTGCTTGC 528
              |   |  | ||
Db        103 GCGGTTTTCGC 93
```

RESULT 56
US-09-252-991A-16005
; Sequence 16005, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16005
; LENGTH: 1530
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16005

Query Match 4.6%; Score 46.2; DB 4; Length 1530;
Best Local Similarity 49.2%; Pred. No. 0.31;
Matches 212; Conservative 0; Mismatches 208; Indels 11; Gaps 3;

```
Qy        103 CGGCGGGGCCCGGCGCGGGGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAAC 162
              | ||  |||   ||  | |      | |  ||  |   |  | | || |    ||
Db        727 CTGCTCGGCGTCGCCGGAGCGTTCGCTGCCGGAGCGCTGGCAGCCGGTGTGGCGGCGGGG 786

Qy        163 GCGAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCT 222
              ||   ||| |||   ||     |    | | | |   | ||  |   |  ||||   ||||
Db        787 GCAGCGGGCGCGCTCGAAGACGCGGCGGGAGCGGCCGGTTTCGCCGCGGCGGGCTTGGCT 846

Qy        223 GCCTGCCTGCTCGCTG--GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGC 280
              ||| || |  |||| |  | ||  ||       ||   |  ||| | ||   ||| |||
Db        847 GCCGGCTTCTTCGCCGCAGGCTTTTTCGCAGCCGGCTTGGCGGCAGGTTTCGCCGCGGGC 906

Qy        281 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGG---GACAGTAGGCCTGCTTGCT 337
               ||   || |  |   | |  ||| ||  ||| |  ||     ||| |||| |    ||
Db        907 TTGGCGGCGGTCGCAGCGGCCGGCTTGGCTGCGGCAGGTTTCGCAGCAGGCTTCGCCGCG 966

Qy        338 CGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCT 397
              || | |||  |||  |  | | ||   |||| | |||   ||| ||  ||   ||
Db        967 GCCTTGGCTGCCGGCTTGGTCGCCGGTTTCGCTGCGGCCTTCGCAGTCGGTTTGGCGGCG 1026

Qy        398 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 457
              ||| ||  |||  ||| |  ||||    |  | ||        | | ||  ||   |||
Db        1027 GGCTTGGCTGCCGGCTTGGCGGCCGCGGTTTTCGCCG------CGGGTTTTTCGCCGCG 1080

Qy        458 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 517
              ||  |    ||  ||| |  |||   |||| |||  | |   |||| |||||  ||||
Db        1081 GGTTTCGCCGCAGGCTTGGCGGCGGCCTTGGCTGCCGGTTTAGCTGCCGGCTTGGCCGCT 1140

Qy        518 GGCCTGCTTGC 528
              |   |  | ||
Db        1141 GCGGTTTTCGC 1151
```

RESULT 57
US-09-780-173A-10/c
; Sequence 10, Application US/09780173A
; Patent No. 6455307
; GENERAL INFORMATION:
; APPLICANT: Robert McKay
; APPLICANT: Susan M. Freier
; APPLICANT: Jacqueline Wyatt
; TITLE OF INVENTION: ANTISENSE MODULATION OF CASEIN KINASE 2-ALPHA PRIME EXPRESSION
; FILE REFERENCE: RTS-0165
; CURRENT APPLICATION NUMBER: US/09/780,173A
; CURRENT FILING DATE: 2001-02-08
; NUMBER OF SEQ ID NOS: 95
; SEQ ID NO 10
; LENGTH: 1877
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (344)...(1396)
US-09-780-173A-10

Query Match 4.6%; Score 46.2; DB 4; Length 1877;
Best Local Similarity 50.6%; Pred. No. 0.32;
Matches 137; Conservative 0; Mismatches 133; Indels 1; Gaps 1;

```
Qy         34 CCCTCCCCCACGCCGCCCCGAAAGGTGGTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCC 93
              |||  |  | |||  |||  |   ||        |||   |||||  || |   |||||| |
Db        371 CCCGACTGCCCGCGGCCGGGCCGGGCATGGCGGGCGGGACCGGGGGGCGCCGCGGGGCGC 312

Qy         94 GGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGACCGAGGGCCCCGGGAGAACGGGGGAT 153
               | | ||  |||||| ||   ||| || ||| | | |  ||| | || ||    |||
Db        311 CGAGGGCGGCGGCGGCGCGGCGCGGGGCGCGGGGCGGCAGGCACGGGAAGCGGCGGGCCG 252

Qy        154 CGGGAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGT-GTGCGTGCTTGTGAGACACC 212
              ||||    |       |||  || |   || | |||||  | |||| |  |  || |
Db        251 CGGGCTGCCCAAGCAAGAGGAGGAGCGCGGCGGCGGCGGCGGGCGTCCGCGCCAGGCCGG 192

Qy        213 GGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGC 272
              |    |||  ||  ||   |  | | |   |  || | ||||   | | | |   |
Db        191 GCCCGCGGGGGCGGGCGGACGGACGGACGCTCGGGCCGGGGGGCGCGGAAGGCGCCGGCC 132

Qy        273 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGG 303
                || |||| |    ||  ||   | | |||
Db        131 GAGCCGGCCCGGGGCCTTCGGCTGCCGGAGG 101
```

RESULT 58
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.

```
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             4.6%;  Score 46.2;  DB 3;  Length 4403765;
  Best Local Similarity   50.7%;  Pred. No. 0.65;
  Matches  111;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy          8 CGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCTCCG 67
              |||   |  | ||  |  | |     ||  |   | || || ||| |  |  | ||
Db    2918217 CGGATGGCTGCTCGGCAACGGTGGCGCCGGCGGTCCCGGCGGCCCAACGGACGTTCCTGC 2918276

Qy         68 CGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGA 127
              || | | ||  | |  | ||||||| | |||| | |  |   |  |||| ||| || |
Db    2918277 CGGCACAGGTGGAGCCGGCGGGGCCGGCGGCGACGCCCCATTGATCGGCTGGGGCGGCAA 2918336

Qy        128 CCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGAC 187
              | | |||||| | |    ||  |  |  |||||  | | ||   ||||       ||  |
Db    2918337 CGGCGGGCCCGGCGGTTTCGCTGCTTTTGGAAACGGTGGGGCCGGCGGCAACGGCGGCGC 2918396

Qy        188 GGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCT 226
                 || ||  |||  | | |  |||   |||| || |
Db    2918397 CAGCGGTTCGCTCTTTGGCGTCGGCGGCGCCGGCGGCGT 2918435


RESULT 59
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             4.6%;  Score 46.2;  DB 3;  Length 4411529;
  Best Local Similarity   50.7%;  Pred. No. 0.65;
  Matches  111;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy          8 CGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCTCCG 67
              |||   |  | ||  |  | |     ||  |   | || || ||| |  |  | ||
Db    2922109 CGGATGGCTGCTCGGCAACGGTGGCGCCGGCGGTCCCGGCGGCCCAACGGACGTTCCTGC 2922168

Qy         68 CGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGA 127
              || | | ||  | |  | ||||||| | |||| | |  |   |  |||| ||| || |
Db    2922169 CGGCACAGGTGGAGCCGGCGGGGCCGGCGGCGACGCCCCATTGATCGGCTGGGGCGGCAA 2922228

Qy        128 CCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGAC 187
              | | |||||| | |    ||  |  |  |||||  | | ||   ||||       ||  |
Db    2922229 CGGCGGGCCCGGCGGTTTCGCTGCTTTTGGAAACGGTGGGGCCGGCGGCAACGGCGGCGC 2922288

Qy        188 GGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCT 226
                 || ||  |||  | | |  |||   |||| || |
Db    2922289 CAGCGGTTCGCTCTTTGGCGTCGGCGGCGCCGGCGGCGT 2922327


RESULT 60
US-09-165-264-12
; Sequence 12, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
;  APPLICANT:  Vinayagamoorthy, Thuraiayah
;  TITLE OF INVENTION:  Multi-Loci Genomic Analysis
;  FILE REFERENCE:  44747
;  CURRENT APPLICATION NUMBER: US/09/165,264
;  CURRENT FILING DATE:  1998-10-01
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 12
;   LENGTH: 318
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-12

  Query Match             4.6%;  Score 45.8;  DB 3;  Length 318;
  Best Local Similarity   56.2%;  Pred. No. 0.32;
  Matches   86;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy         57 GGTGGTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGC 116
              || ||     | |  |  ||| ||||||  ||||   ||||  |    || ||||   ||
Db        164 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 223

Qy        117 GCGGGGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGG 176
              | |||| | |   | |||    ||| |   |||||   |||     | | |||| | |||
Db        224 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 283

Qy        177 GGACAGGGGACGGCGTGTGCGTGCTTGTGAGAC 209
              ||   ||||  || | ||  |||  ||| || |
Db        284 GGGGGGGGGGGGGGGGGTTGGTGTGTGTAAGCC 316


RESULT 61
US-09-759-359A-3
; Sequence 3, Application US/09759359A
; Patent No. 6492153
; GENERAL INFORMATION:
;  APPLICANT: ABU-THREIDEH, Jane et al
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001043
;  CURRENT APPLICATION NUMBER: US/09/759,359A
;  CURRENT FILING DATE:  2001-01-16
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 90541
;   TYPE: DNA
;   ORGANISM: Human
US-09-759-359A-3

  Query Match             4.6%;  Score 45.8;  DB 4;  Length 90541;
  Best Local Similarity   57.2%;  Pred. No. 0.54;
  Matches   83;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy        393 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 452
```

```
                || ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | | | ||
Db      34807 TCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCCTTTTTCT 34866

Qy        453 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 512
              |       || ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| |
Db      34867 TTTTCTTTCTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC 34926

Qy        513 TCGCTGGCCTGCTTGCTTGCTCGCT 537
              |  ||  | | ||| ||| ||  ||
Db      34927 TTTCTTTCTTTCTTCCTTTCTTTCT 34951


RESULT 62
US-09-165-264-8
; Sequence 8, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
;  APPLICANT:  Vinayagamoorthy, Thuraiayah
;  TITLE OF INVENTION:  Multi-Loci Genomic Analysis
;  FILE REFERENCE:  44747
;  CURRENT APPLICATION NUMBER: US/09/165,264
;  CURRENT FILING DATE:  1998-10-01
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 319
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-8

  Query Match             4.6%;  Score 45.6;  DB 3;  Length 319;
  Best Local Similarity  47.8%;  Pred. No. 0.35;
  Matches  132;  Conservative    0;  Mismatches  144;  Indels    0;  Gaps    0;

Qy         57 GGTGGTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGC 116
              || ||     | |  |  ||| ||||||  ||||   |||| |    || ||||   ||
Db         30 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 89

Qy        117 GCGGGGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGG 176
              | |||| | |   | |||    ||| |   |||||   |||     | | |||| | |||
Db         90 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 149

Qy        177 GGACAGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGC 236
              ||   ||||  || | | | | |   | | |     |||   ||  |   |   |   |
Db        150 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 209

Qy        237 TGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 296
               ||   |   | | ||||   |  ||   |   |   |  ||   |   |  | ||||
Db        210 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 269

Qy        297 CAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGC 332
                |  ||   |   | | ||||   |  || ||||
Db        270 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGTCTGC 305


RESULT 63
US-09-019-095A-7/c
; Sequence 7, Application US/09019095A
; Patent No. 6287858
; GENERAL INFORMATION:
;  APPLICANT:  D'Andrea, Alan D.
;  APPLICANT:  Zhu, Yuan
;  TITLE OF INVENTION:  Deubiquitinating Enzymes That Regulate
;  TITLE OF INVENTION:  Cell Growth
;  FILE REFERENCE:  DFCI-435p2A2
;  CURRENT APPLICATION NUMBER: US/09/019,095A
;  CURRENT FILING DATE:  1998-02-05
;  PRIOR APPLICATION NUMBER:  PCT/US96/12884
;  PRIOR FILING DATE:  1996-08-07
;  PRIOR APPLICATION NUMBER:  US 60/002,066
;  PRIOR FILING DATE:  1995-08-09
;  PRIOR APPLICATION NUMBER:  US 60/019,787
;  PRIOR FILING DATE:  1996-06-14
;  NUMBER OF SEQ ID NOS:  51
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 1559
;   TYPE: DNA
;   ORGANISM: Murine
US-09-019-095A-7

  Query Match             4.6%;  Score 45.6;  DB 3;  Length 1559;
  Best Local Similarity  55.8%;  Pred. No. 0.41;
  Matches   87;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy        394 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 453
              | ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||
Db       1137 CCCTTTCTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 1078

Qy        454 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 513
              ||  | |  |  ||  ||| ||      ||| ||  ||| ||  |   ||  ||| ||
Db       1077 TCCTTTCTTTATCTCTCTCTCTCTTTCTCTTTCTCTCTCTCTATCTCTCTCTCTTCCT 1018

Qy        514 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              ||  | | | | ||| ||  ||  | | ||| ||
Db       1017 TCCTTCCGTCCGTCCTTCCTTCCTTCCTTCCTTCCT 982


RESULT 64
US-09-499-522-1/c
; Sequence 1, Application US/09499522
; Patent No. 6479238
; GENERAL INFORMATION:
;  APPLICANT: Blumenfeld, Marta
;  APPLICANT: Bougueleret, Lydie
;  APPLICANT: Bihain, Bernard
;  TITLE OF INVENTION: POLYMORPHIC MARKERS OF THE LSR GENE
;  FILE REFERENCE: GENSET.053AUS
;  CURRENT APPLICATION NUMBER: US/09/499,522
;  CURRENT FILING DATE: 2000-02-10
;  EARLIER APPLICATION NUMBER: US 60/119,592
;  EARLIER FILING DATE: 1999-02-10
;  EARLIER APPLICATION NUMBER: US 60/144,784
;  EARLIER FILING DATE: 1999-07-20
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: Patent.pm
; SEQ ID NO 1
;   LENGTH: 23187
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 2001..2356
;   OTHER INFORMATION: exon1
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 3540..3884
;   OTHER INFORMATION: exon2
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 12163..12282
;   OTHER INFORMATION: exon3
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 15144..15200
;   OTHER INFORMATION: exon4
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 15765..15911
;   OTHER INFORMATION: exon5
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 19579..19752
;   OTHER INFORMATION: exon6
```

```
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 19899..19958
;   OTHER INFORMATION: exon7
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 20056..20187
;   OTHER INFORMATION: exon8
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 20329..20957
;   OTHER INFORMATION: exon9
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 21047..21187
;   OTHER INFORMATION: exon10
;   FEATURE:
;   NAME/KEY: polyA_signal
;   LOCATION: 21168..21173
;   OTHER INFORMATION: AATAAA
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..2000
;   OTHER INFORMATION: potential 5'regulatory region
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 22324..23187
;   OTHER INFORMATION: homology with USF2 gene in ref: embl Y07661
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 523..544
;   OTHER INFORMATION: upstream amplification primer 17-2
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 1047..1068
;   OTHER INFORMATION: downstream amplification primer 17-2 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 946..963
;   OTHER INFORMATION: upstream amplification primer 99-4576
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 1385..1402
;   OTHER INFORMATION: downstream amplification primer 99-4576 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 1096..1115
;   OTHER INFORMATION: upstream amplification primer 9-19
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 1616..1635
;   OTHER INFORMATION: downstream amplification primer 9-19 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 1602..1621
;   OTHER INFORMATION: upstream amplification primer 9-20
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 2074..2093
;   OTHER INFORMATION: downstream amplification primer 9-20 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 2036..2053
;   OTHER INFORMATION: upstream amplification primer 99-4557
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 2563..2580
;   OTHER INFORMATION: downstream amplification primer 99-4557 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 2084..2102
;   OTHER INFORMATION: upstream amplification primer 9-1
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 2483..2500
;   OTHER INFORMATION: downstream amplification primer 9-1 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 2470..2489
;   OTHER INFORMATION: upstream amplification primer 9-21 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 2062..2081
;   OTHER INFORMATION: downstream amplification primer 9-21
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 3455..3474
;   OTHER INFORMATION: upstream amplification primer 9-3
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 3882..3901
;   OTHER INFORMATION: downstream amplification primer 9-3 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 3775..3792
;   OTHER INFORMATION: upstream amplification primer 99-4558
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 4336..4356
;   OTHER INFORMATION: downstream amplification primer 99-4558 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 4902..4920
;   OTHER INFORMATION: upstream amplification primer 99-14419 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 4444..4463
;   OTHER INFORMATION: downstream amplification primer 99-14419
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 6638..6655
;   OTHER INFORMATION: upstream amplification primer 99-4577
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 7072..7089
;   OTHER INFORMATION: downstream amplification primer 99-4577 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 7995..8012
;   OTHER INFORMATION: upstream amplification primer 99-4559
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 8576..8593
;   OTHER INFORMATION: downstream amplification primer 99-4559 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 9622..9639
;   OTHER INFORMATION: upstream amplification primer 99-3148
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 10023..10040
;   OTHER INFORMATION: downstream amplification primer 99-3148 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 9964..9981
;   OTHER INFORMATION: upstream amplification primer 99-4560
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 10546..10563
;   OTHER INFORMATION: downstream amplification primer 99-4560 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 10996..11015
;   OTHER INFORMATION: upstream amplification primer 99-14411 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
```

```
;   LOCATION: 10492..10512
;   OTHER INFORMATION: downstream amplification primer 99-14411
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 11972..11990
;   OTHER INFORMATION: upstream amplification primer 99-4561
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 12481..12501
;   OTHER INFORMATION: downstream amplification primer 99-4561 , complement

  Query Match             4.5%;  Score 45.4;  DB 4;  Length 23187;
  Best Local Similarity   51.8%;  Pred. No. 0.58;
  Matches  100;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy          1 CGCCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTG 60
              ||  ||| ||    | || |     | ||||     ||   |  | || |   |   | |
Db      22428 CGGGGCCGGGGCGCGGGGGCACGGCCGGGCCGGGGATCTTACCTGGCGGCGGCAGCAGCG 22369

Qy         61 GTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGG 120
              ||  ||| | |  |||||    | || ||| || |   | |   | ||||  ||| |  |
Db      22368 GTGGCCGAGGCAGCGGGATCCAGACCCGGGTCCAGCATGTCCATGGGGGGGGCGGGGGCG 22309

Qy        121 GGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGAC 180
              || | | |   |||||| ||| | | ||| |    ||  |  |   | || | |||||
Db      22308 GGGGGGGCGCGGGGCCCGGGGGGGAGGGGAGGGGAGGGGANGGANGGAGGGGAGGGGCG 22249

Qy        181 AGGGGACGGCGTG 193
              | || ||| | |
Db      22248 GGCGGCCGGGGGG 22236


RESULT 65
US-08-585-593A-42/c
; Sequence 42, Application US/08585593A
; Patent No. 6503706
;  GENERAL INFORMATION:
;    APPLICANT:  ABKEN, Hinrich J
;    APPLICANT:  ALBERT, Winfried
;    APPLICANT:  JUNGFER, Herbert
;    TITLE OF INVENTION:  METHOD OF IDENTIFYING HUMAN AND ANIMAL
;    TITLE OF INVENTION:  CELLS CAPABLE OF UNLIMITED PROLIFERATION OR TUMOR
;    TITLE OF INVENTION:  FORMATION
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nikaido, Marmelstein, Murray & Oram LLP
;      STREET:  655 Fifteenth Street N.W. Suite 330
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-5701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/585,593A
;      FILING DATE:  16-JAN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP94/02307
;      FILING DATE:  13-JUL-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P 43 23 727.4
;      FILING DATE:  15-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murray, Robert B.
;      REGISTRATION NUMBER:  22,980
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)638-5000
;      TELEFAX:  (202)638-4810
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  275 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-585-593A-42

  Query Match             4.5%;  Score 45;  DB 4;  Length 275;
  Best Local Similarity   56.4%;  Pred. No. 0.46;
  Matches   84;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy        389 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 448
              || ||   |   || |||  |  ||| ||  |   ||  ||  ||| ||  | | ||| |
Db        246 CTTCTTCTTCTTCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTTCTTTCTTTC 187

Qy        449 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 508
              || ||  ||  | | ||| ||  ||  ||  |   ||| ||| ||  ||  | | ||| |
Db        186 TTTCTTTCTTTCTTTCTTCCTGTCTGCCTACCTGCCTTCCTTCCTTCCTTCCTTCCTTCC 127

Qy        509 TTGCTCGCTGGCCTGCTTGCTTGCTCGCT 537
              || ||  ||  | | ||| ||| ||  ||
Db        126 TTCCTTCCTTCCTTCCTTCCTTCCTTTCT 98


RESULT 66
US-09-165-264-14
; Sequence 14, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
;  APPLICANT:  Vinayagamoorthy, Thuraiayah
;  TITLE OF INVENTION:  Multi-Loci Genomic Analysis
;  FILE REFERENCE:  44747
;  CURRENT APPLICATION NUMBER: US/09/165,264
;  CURRENT FILING DATE:  1998-10-01
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 14
;   LENGTH: 320
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-14

  Query Match             4.5%;  Score 45;  DB 3;  Length 320;
  Best Local Similarity   46.4%;  Pred. No. 0.47;
  Matches  147;  Conservative    0;  Mismatches  170;  Indels    0;  Gaps    0;

Qy         75 GGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGACCGAGGG 134
              ||| ||||||  ||||   |||| |    || ||||   || | |||| | |   | |||
Db          1 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 60

Qy        135 CCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGTGT 194
                 ||| |   |||||   |||     | | |||| | |||||   ||||  || | |
Db         61 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120

Qy        195 GCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGG 254
              | | |   | | |     |||   ||  |   |   |   |  ||   |   | | |||
Db        121 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 180

Qy        255 GACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTG 314
              |   |  ||   |   |   |  ||   |   | | ||||   |  ||   |   | |
Db        181 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 240

Qy        315 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTG 374
               ||||   |  ||   |   |   |  ||   |   | | ||||   |  |||  | |
Db        241 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 300

Qy        375 GCTTGCTAGTAGGGCTG 391
                |  |  |||  || |
Db        301 TGTGCCATGTACAGCAG 317
```

```
RESULT 67
US-08-306-691B-14
; Sequence 14, Application US/08306691B
; Patent No. 5734039
;  GENERAL INFORMATION:
;    APPLICANT:  Calabretta, Bruno
;    APPLICANT:  Skorski, Tomasz
;    TITLE OF INVENTION:  ANTISENSE
;    TITLE OF INVENTION:  OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
;    NUMBER OF SEQUENCES:  55
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seidel, Gonda, Lavorgna & Monaco, P.C.
;      STREET:  Two Penn Center, Suite 1800
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  U.S.A.
;      ZIP:  19102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 720 Kb
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/306,691B
;      FILING DATE:  September 15, 1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monaco, Daniel A.
;      REGISTRATION NUMBER:  30,480
;      REFERENCE/DOCKET NUMBER:  8321-8
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-8383
;      TELEFAX:  (215) 568-5549
;      TELEX:  No. 5734039e
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6453 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-306-691B-14

  Query Match             4.5%;  Score 45;  DB 1;  Length 6453;
  Best Local Similarity   49.8%;  Pred. No. 0.62;
  Matches  114;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy          2 GCCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGG 61
              || | | || || |       |   || ||||| | |||||| || ||  | | | | |
Db        583 GCGGGCAGGGATCGGCCGGGCCGCGCGCCCTCCTCGCCCCCAGGCGGCAGCAATACGCGC 642

Qy         62 TCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGG 121
                 || ||||  ||   |||| | | || |   ||| |   |  ||   |||   ||
Db        643 GGCGCGGGCCGGGGGCGCGGGGCCGGCGGGCGTAAGCGGCGGCGGCGGCGGCGGGTGGGT 702

Qy        122 GCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACA 181
              | |     | ||| |||| | | || || || |||     | |   |  ||||| |
Db        703 GGGGCCGGGCGGGGCCCGCGGGCACAGGTGAGCGGGCGTCGGGGGCTGCGGCGGGCGGGG 762

Qy        182 GGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCT 230
              |       |      | || || | ||  |||||  |  |     ||||
Db        763 GCCCCTTCCTCCCTGGGGCCTGCGGGAATCCGGGCCCCACCCGTGGCCT 811


RESULT 68
US-09-209-668-10
; Sequence 10, Application US/09209668A
; Patent No. 6114517
; GENERAL INFORMATION:
;  APPLICANT: Monia, Brett P.
;  APPLICANT: Xu, Xiaoxing S.
;  TITLE OF INVENTION: METHODS OF MODULATING TUMOR NECROSIS FACTOR
;  TITLE OF INVENTION: alpha-INDUCED EXPRESSION OF CELL ADHESION MOLECULES
;  FILE REFERENCE: ISPH-0336
;  CURRENT APPLICATION NUMBER: US/09/209,668A
;  CURRENT FILING DATE: 1998-12-10
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 6453
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1664)..(1774)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2042)..(2220)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2374)..(2533)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (3231)..(3350)
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: V00574/Genbank
;   DATABASE ENTRY DATE: 1991-01-03
US-09-209-668-10

  Query Match             4.5%;  Score 45;  DB 3;  Length 6453;
  Best Local Similarity   49.8%;  Pred. No. 0.62;
  Matches  114;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy          2 GCCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGG 61
              || | | || || |       |   || ||||| | |||||| || ||  | | | | |
Db        583 GCGGGCAGGGATCGGCCGGGCCGCGCGCCCTCCTCGCCCCCAGGCGGCAGCAATACGCGC 642

Qy         62 TCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGG 121
                 || ||||  ||   |||| | | || |   ||| |   |  ||   |||   ||
Db        643 GGCGCGGGCCGGGGGCGCGGGGCCGGCGGGCGTAAGCGGCGGCGGCGGCGGCGGGTGGGT 702

Qy        122 GCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACA 181
              | |     | ||| |||| | | || || || |||     | |   |  ||||| |
Db        703 GGGGCCGGGCGGGGCCCGCGGGCACAGGTGAGCGGGCGTCGGGGGCTGCGGCGGGCGGGG 762

Qy        182 GGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCT 230
              |       |      | || || | ||  |||||  |  |     ||||
Db        763 GCCCCTTCCTCCCTGGGGCCTGCGGGAATCCGGGCCCCACCCGTGGCCT 811


RESULT 69
US-09-356-952-8
; Sequence 8, Application US/09356952
; Patent No. 6117663
; GENERAL INFORMATION:
;  APPLICANT: Boriack-Sjodin, Ann
;  APPLICANT: Margarit, S. M.
;  APPLICANT: Bor-Sogi, Dafna
;  APPLICANT: Cole, Philip
;  APPLICANT: Kuriyan, John
;  TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
;  TITLE OF INVENTION: THEREOF
;  FILE REFERENCE: 600-1-228N
;  CURRENT APPLICATION NUMBER: US/09/356,952
;  CURRENT FILING DATE: 1999-07-19
;  EARLIER APPLICATION NUMBER: 60/093,631
;  EARLIER FILING DATE: 1998-07-21
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 6453
;   TYPE: DNA
;   ORGANISM: Homo sapiens
```

```
US-09-356-952-8

  Query Match             4.5%;  Score 45;  DB 3;  Length 6453;
  Best Local Similarity   49.8%;  Pred. No. 0.62;
  Matches  114;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy          2 GCCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGG 61
              || | | || || |        |   || ||||| | |||||| || ||  | | | | |
Db        583 GCGGGCAGGGATCGGCCGGGCCGCGCGCCCTCCTCGCCCCCAGGCGGCAGCAATACGCGC 642

Qy         62 TCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGG 121
                 || ||||  ||   |||| | | || |    ||| |   |  ||   |||   ||
Db        643 GGCGCGGGCCGGGGGCGCGGGGCCGGCGGGCGTAAGCGGCGGCGGCGGCGGCGGGTGGGT 702

Qy        122 GCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACA 181
              | |     | ||| |||| |.| || || || |||     | |   |  ||||| |
Db        703 GGGGCCGGGCGGGGCCCGCGGGCACAGGTGAGCGGGCGTCGGGGGCTGCGGCGGGCGGGG 762

Qy        182 GGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCT 230
              |       |      | || || | ||  |||||  |  |     ||||
Db        763 GCCCCTTCCTCCCTGGGGCCTGCGGGAATCCGGGCCCCACCCGTGGCCT 811


RESULT 70
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             4.5%;  Score 44.8;  DB 3;  Length 4403765;
  Best Local Similarity   55.0%;  Pred. No. 1.2;
  Matches   88;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy         68 CGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGA 127
              || |||| |  | ||||| || |     ||||   | ||   |  |||    ||  ||||
Db    2422605 CGGCGCCAGCGGTGGGGCGGGTGGTGCCGGCGGGGCCGCCCTGATCGGTAATGGCGGCGA 2422546

Qy        128 CCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGAC 187
              | | |||| | | |  ||||| ||    || ||  |||  |||  ||| ||    ||
Db    2422545 CGGGGGGCACGGCGGCAACGGTGGCCACGGCAACAGCGGTGGGGCCGGCGGCGCCGGTGG 2422486

Qy        188 GGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTG 227
              |||  |||  |  |    | | |   ||| ||| | | |||
Db    2422485 GGCCGGTGGGGCCGGTGGGGCCGGTGGGCACGTCGGGCTG 2422446


RESULT 71
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             4.5%;  Score 44.8;  DB 3;  Length 4411529;
  Best Local Similarity   55.0%;  Pred. No. 1.2;
  Matches   88;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy         68 CGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGA 127
              || |||| |  | ||||| || |     ||||   | ||   |  |||    ||  ||||
Db    2423946 CGGCGCCAGCGGTGGGGCGGGTGGTGCCGGCGGGGCCGCCCTGATCGGTAATGGCGGCGA 2423887

Qy        128 CCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGAC 187
              | | |||| | | |  ||||| ||    || ||  |||  |||  ||| ||    ||
Db    2423886 CGGGGGGCACGGCGGCAACGGTGGCCACGGCAACAGCGGTGGGGCCGGCGGCGCCGGTGG 2423827

Qy        188 GGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTG 227
              |||  |||  |  |    | | |   ||| ||| | | |||
Db    2423826 GGCCGGTGGGGCCGGTGGGGCCGGTGGGCACGTCGGGCTG 2423787


RESULT 72
US-09-056-556-165
; Sequence 165, Application US/09056556
; Patent No. 6350456
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Skeiky, Yasir A.W.
;    APPLICANT:  Dillon, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE PREVENTION AND           TRE
;    NUMBER OF SEQUENCES:  241
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/056,556
;      FILING DATE:  07-APR-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.457
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  165:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  227 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-056-556-165
```

```
  Query Match             4.4%;  Score 44.2;  DB 4;  Length 227;
  Best Local Similarity   52.4%;  Pred. No. 0.66;
  Matches   97;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy          5 GCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCT 64
              | ||||||||     ||||  || ||  |    | |||| | || |    |||| ||
Db         20 GGCCGGCATGAACAGCCTCGACCCGCTGCTAGCCGCCCAAGACGGCGGCCAAGGCGGCAC 79

Qy         65 CCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCG 124
              | ||| | ||||  |    |||| ||| | ||| ||         |  |||| |   |
Db         80 CGGCGGCACCGGCGGCAACGCCGGCGCCGGCGGCACCAGCTTCACCCAAGGCGCCGACGG 139

Qy        125 CGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGG 184
              | ||   ||   | | |   |||| ||    ||      ||  || | ||| ||| | ||
Db        140 CAACGCCGGCAACGGCGGTGACGGCGGGGTCGGCGGCAACGGCGGAAACGGCGGAAACGG 199

Qy        185 GACGG 189
                | |
Db        200 CGCAG 204


RESULT 73
US-09-072-596-160
; Sequence 160, Application US/09072596
; Patent No. 6458366
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Skeiky, Yasir A.W.
;    APPLICANT:  Dillon, Davin C.
;    APPLICANT:  Campos-Neto, Antonia
;    APPLICANT:  Houghton, Raymond
;    APPLICANT:  Vedvick, Thomas S.
;    APPLICANT:  Twardzik, Daniel R.
;    APPLICANT:  Lodes, Michael J.
;    APPLICANT:  Hendrickson, Ronald C.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR DIAGNOSIS OF    TUBERCULOSIS
;    NUMBER OF SEQUENCES:  350
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/072,596
;      FILING DATE:  05-MAY-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.417C9
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  160:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  227 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-072-596-160

  Query Match             4.4%;  Score 44.2;  DB 4;  Length 227;
  Best Local Similarity   52.4%;  Pred. No. 0.66;
  Matches   97;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy          5 GCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCT 64
              | ||||||||     ||||  || ||  |    | |||| | || |    |||| ||
Db         20 GGCCGGCATGAACAGCCTCGACCCGCTGCTAGCCGCCCAAGACGGCGGCCAAGGCGGCAC 79

Qy         65 CCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCG 124
              | ||| | ||||  |    |||| ||| | ||| ||         |  |||| |   |
Db         80 CGGCGGCACCGGCGGCAACGCCGGCGCCGGCGGCACCAGCTTCACCCAAGGCGCCGACGG 139

Qy        125 CGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGG 184
              | ||   ||   | | |   |||| ||    ||      ||  || | ||| ||| | ||
Db        140 CAACGCCGGCAACGGCGGTGACGGCGGGGTCGGCGGCAACGGCGGAAACGGCGGAAACGG 199

Qy        185 GACGG 189
                | |
Db        200 CGCAG 204


RESULT 74
US-09-056-556-179
; Sequence 179, Application US/09056556
; Patent No. 6350456
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Skeiky, Yasir A.W.
;    APPLICANT:  Dillon, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE PREVENTION AND           TRE
;    NUMBER OF SEQUENCES:  241
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/056,556
;      FILING DATE:  07-APR-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.457
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  179:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  400 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-056-556-179

  Query Match             4.4%;  Score 44.2;  DB 4;  Length 400;
  Best Local Similarity   52.4%;  Pred. No. 0.69;
  Matches   97;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy          5 GCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCT 64
              | ||||||||     ||||  || ||  |    | |||| | || |    |||| ||
Db         20 GGCCGGCATGAACAGCCTCGACCCGCTGCTAGCCGCCCAAGACGGCGGCCAAGGCGGCAC 79

Qy         65 CCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCG 124
              | ||| | ||||  |    |||| ||| | ||| ||         |  |||| |   |
Db         80 CGGCGGCACCGGCGGCAACGCCGGCGCCGGCGGCACCAGCTTCACCCAAGGCGCCGACGG 139

Qy        125 CGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGG 184
              | ||   ||   | | |   |||| ||    ||      ||  || | ||| ||| | ||
Db        140 CAACGCCGGCAACGGCGGTGACGGCGGGGTCGGCGGCAACGGCGGAAACGGCGGAAACGG 199
```

```
Qy        185 GACGG 189
                | |
Db        200 CGCAG 204


RESULT 75
US-09-072-596-174
; Sequence 174, Application US/09072596
; Patent No. 6458366
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Skeiky, Yasir A.W.
;    APPLICANT:  Dillon, Davin C.
;    APPLICANT:  Campos-Neto, Antonia
;    APPLICANT:  Houghton, Raymond
;    APPLICANT:  Vedvick, Thomas S.
;    APPLICANT:  Twardzik, Daniel R.
;    APPLICANT:  Lodes, Michael J.
;    APPLICANT:  Hendrickson, Ronald C.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR DIAGNOSIS OF    TUBERCULOSIS
;    NUMBER OF SEQUENCES:  350
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/072,596
;      FILING DATE:  05-MAY-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.417C9
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  174:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  400 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-072-596-174

  Query Match             4.4%;  Score 44.2;  DB 4;  Length 400;
  Best Local Similarity   52.4%;  Pred. No. 0.69;
  Matches   97;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy          5 GCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCT 64
              | ||||||||     ||||  || ||  |    | |||| | || |    |||| ||
Db         20 GGCCGGCATGAACAGCCTCGACCCGCTGCTAGCCGCCCAAGACGGCGGCCAAGGCGGCAC 79

Qy         65 CCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCG 124
              | ||| | ||||  |    ||||| ||| | ||| ||          |  ||||| |   |
Db         80 CGGCGGCACCGGCGGCAACGCCGGCGCCGGCGGCACCAGCTTCACCCAAGGCGCCGACGG 139

Qy        125 CGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGG 184
              | ||   ||   | | |   |||| ||    ||      ||  || | ||| ||| | ||
Db        140 CAACGCCGGCAACGGCGGTGACGGCGGGGTCGGCGGCAACGGCGGAAACGGCGGAAACGG 199

Qy        185 GACGG 189
                | |
Db        200 CGCAG 204


RESULT 76
US-09-128-155-17
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
;  APPLICANT: Pan, Yang
;  TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
;  TITLE OF INVENTION: AND USES THEREOF
;  FILE REFERENCE: 09404/052001
;  CURRENT APPLICATION NUMBER: US/09/128,155
;  CURRENT FILING DATE: 1998-08-03
;  EARLIER APPLICATION NUMBER: US 60/091,650
;  EARLIER FILING DATE: 1998-07-02
;  EARLIER APPLICATION NUMBER: US 60/054,646
;  EARLIER FILING DATE: 1997-08-04
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 17
;   LENGTH: 176373
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(176373)
;   OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

  Query Match             4.4%;  Score 44.2;  DB 3;  Length 176373;
  Best Local Similarity   59.1%;  Pred. No. 1.2;
  Matches   94;  Conservative    0;  Mismatches   63;  Indels    2;  Gaps    1;

Qy        393 TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC 452
              ||  |  |'| ||| ||| ||  ||  | | ||| ||| ||  |  || |||||||| |
Db     122289 TCTTTATCTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTTCCTGCTTGCTTGCTTTC 122348

Qy        453 TCGCTGGCCTGCTTGCTTGC--TCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 510
              |  ||  ||| | | | | |  || ||  | | ||| ||   ||  |  | | ||| |||
Db     122349 TTTCTTTCCTTCCTTCCTTCTTTCTCTCCCTTTCTTTCTCTTTCTTTCTCTTTCTTTCTT 122408

Qy        511 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
                ||  ||  | | ||| ||| ||  ||  | | ||| ||
Db     122409 TCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 122447


RESULT 77
US-08-770-379-16
; Sequence 16, Application US/08770379
; Patent No. 5849564
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
;    TITLE OF INVENTION:  HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,379
;      FILING DATE:
```

```
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  52342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  801 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
US-08-770-379-16

  Query Match             4.4%;  Score 44;  DB 2;  Length 801;
  Best Local Similarity   54.3%;  Pred. No. 0.81;
  Matches   89;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy         14 GGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCTCCGCGCCGC 73
              | ||| |  |||| ||   |    | | | || || | |    |  || |   |  | |
Db         94 GCCGGCCGCCCGCGGGGGGCAGGGCGCGCCCGGCGGCTCCCTCGCGGGGCGGGGGACGGG 153

Qy         74 CGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGACCGAGG 133
              | | |||| |||||| ||| | ||||| |||| | |   ||| ||| | |  |||
Db        154 GGAGGGGGGCGCCGGGCCCCCGCGCGCCGCGGCAGCGGAGCGCGAGGGCCCCCGCCGGCC 213

Qy        134 GCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGG 177
              |||   || |     ||   ||||    || |    ||||||||
Db        214 GCCAGCGGCGGCGCAGGCCCCGGGGGCCCGAGCCCCGAGCGGGG 257


RESULT 78
US-08-757-669A-16
; Sequence 16, Application US/08757669A
; Patent No. 6183751
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
;    TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/757,669A
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  45185-F
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  801 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
US-08-757-669A-16

  Query Match             4.4%;  Score 44;  DB 3;  Length 801;
  Best Local Similarity   54.3%;  Pred. No. 0.81;
  Matches   89;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy         14 GGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCTCCGCGCCGC 73
              | ||| |  |||| ||   |    | | | || || | |    |  || |   |  | |
Db         94 GCCGGCCGCCCGCGGGGGGCAGGGCGCGCCCGGCGGCTCCCTCGCGGGGCGGGGGACGGG 153

Qy         74 CGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGACCGAGG 133
              | | |||| |||||| ||| | ||||| |||| | |   ||| ||| | |  |||
Db        154 GGAGGGGGGCGCCGGGCCCCCGCGCGCCGCGGCAGCGGAGCGCGAGGGCCCCCGCCGGCC 213

Qy        134 GCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGG 177
              |||   || |     ||   ||||    || |    ||||||||
Db        214 GCCAGCGGCGGCGCAGGCCCCGGGGGCCCGAGCCCCGAGCGGGG 257


RESULT 79
US-09-230-371A-16
; Sequence 16, Application US/09230371A
; Patent No. 6348586
; GENERAL INFORMATION:
;  APPLICANT:  Chang, Yuan
;  APPLICANT:  Bohenzky, Roy A
;  APPLICANT:  Russo, James J
;  APPLICANT:  Edelman, Isidore S
;  APPLICANT:  Moore, Patrick S
;  TITLE OF INVENTION:  UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE:  45185-G-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/230,371A
;  CURRENT FILING DATE:  1999-11-17
;  PRIOR APPLICATION NUMBER:  PCT/US97/13346
;  PRIOR FILING DATE:  1997-07-22
;  NUMBER OF SEQ ID NOS:  30
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 16
;   LENGTH: 801
;   TYPE: DNA
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-16

  Query Match             4.4%;  Score 44;  DB 4;  Length 801;
  Best Local Similarity   54.3%;  Pred. No. 0.81;
  Matches   89;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy         14 GGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCTCCGCGCCGC 73
              | ||| |  |||| ||   |    | | | || || | |    |  || |   |  | |
Db         94 GCCGGCCGCCCGCGGGGGGCAGGGCGCGCCCGGCGGCTCCCTCGCGGGGCGGGGGACGGG 153

Qy         74 CGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGACCGAGG 133
              | | |||| |||||| ||| | ||||| |||| | |   ||| ||| | |  |||
Db        154 GGAGGGGGGCGCCGGGCCCCCGCGCGCCGCGGCAGCGGAGCGCGAGGGCCCCCGCCGGCC 213

Qy        134 GCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGG 177
              |||   || |     ||   ||||    || |    ||||||||
Db        214 GCCAGCGGCGGCGCAGGCCCCGGGGGCCCGAGCCCCGAGCGGGG 257


RESULT 80
US-08-458-568A-11/c
; Sequence 11, Application US/08458568A
; Patent No. 5821339
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Schaffer, Priscilla A.
;    APPLICANT:  Yeh, Lily
;    TITLE OF INVENTION:  Compositions and Methods for Treatment of Herpesvirus
;    TITLE OF INVENTION:  Infections
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339ris
;      STREET:  One Liberty Place, 46th floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,568A
;      FILING DATE:  02-JUNE-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/065,146
;      FILING DATE:  05-MAY-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Leary Ph.D., Kathryn R.
;      REGISTRATION NUMBER:  36,317
;      REFERENCE/DOCKET NUMBER:  DFCI-0029
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12001 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Herpes simplex virus
;      STRAIN:  Herpes Simplex Virus Type 1
US-08-458-568A-11

  Query Match             4.4%;  Score 44;  DB 1;  Length 12001;
  Best Local Similarity   51.7%;  Pred. No. 1;
  Matches  124;  Conservative    0;  Mismatches  115;  Indels    1;  Gaps    1;

Qy          1 CGCCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTG 60
              ||||||| ||   ||| |||| ||| || || || | ||| |  |  ||  |||
Db       4741 CGCCGCCTGGGGCGGCTGTCCGCCG-CGCCCGCCTCCCCCGCGGGGGGCGACGACCCCGA 4683

Qy         61 GTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGG 120
                |  | |||| |   || |  | |   | | | ||||||| || |  |||| | ||||
Db       4682 CGCCGCCCGCCACGCCGACGCCGACGACGACGCCGGGCGCCGCGCCCAGGCCGGCCGCGT 4623

Qy        121 GGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGAC 180
              |||       | |      | |      || ||||||  | |  |||  |  ||| |
Db       4622 GGCCGTCGAGTGCCTGGCCGCCTGCCGCGGGATCCTGGAGGCGCTGGCCGAGGGCTTCGA 4563

Qy        181 AGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGC 240
               || |||   | |  ||| |  | |     | |||  ||||   |   |  | ||| | |
Db       4562 CGGCGACCTGGCGGCCGTCCCGGGGCTGGCCGGGGCCCGGCCCGCCAGCCCCCCGCGGCC 4503


RESULT 81
US-08-984-709A-52
; Sequence 52, Application US/08984709A
; Patent No. 6320032
;  GENERAL INFORMATION:
;    APPLICANT:  Williams, Mark E.
;    APPLICANT:  Stauderman, Kenneth A.
;    APPLICANT:  Harpold, Michael M.
;    TITLE OF INVENTION:  HUMAN CALCIUM CHANNEL COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Heller Ehrman White & McAuliffe
;      STREET:  4250 Executive Square, Suite 700
;      CITY:  La Jolla
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/984,709A
;      FILING DATE:  02-DEC-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L.
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  24735-9815 (formerly 6362-9815)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 450-8400
;      TELEFAX:  (619) 587-5360
;  INFORMATION FOR SEQ ID NO:  52:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1413 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-984-709A-52

  Query Match             4.4%;  Score 43.8;  DB 4;  Length 1413;
  Best Local Similarity   49.4%;  Pred. No. 0.93;
  Matches  114;  Conservative    0;  Mismatches  117;  Indels    0;  Gaps    0;

Qy          1 CGCCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTG 60
              | | |||||        | ||  ||||   | ||||  || ||| |  || |  | |
Db         35 CCCGGCCCGCGCCCCGCGCCCCGCGCCCCGCGCCCCGGCCTCACCCGTCCGCTCAGCGGC 94

Qy         61 GTCTCCGCGCCGCCGGGAGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGG 120
                ||  |||  ||||| |   |  ||||  |||   | ||     || ||  |||| |  |
Db         95 CTCCACGCCGCGCCGAGGCCGCCGCCGTCGCCTCCGCCGGGCGAGCCGGAGCCGGAGTCG 154

Qy        121 GGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGAC 180
              ||      |  ||  || ||  |   ||||   ||||      |   | || || |||| |
Db        155 AGCCGCGGCCGGGAGCCGGGCGGGCTGGGGACGCGGGCCGGGGGCGGAGGCGCTGGGGGC 214

Qy        181 AGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTG 231
               |||| ||| |   |   |  | | |   |  |||      |||  || |
Db        215 CGGGGCCGGGGCCGGGCGCCGAGCGGGGTCCGCGGTGACCGCGCCGCCCGG 265


RESULT 82
US-08-485-284A-1/c
; Sequence 1, Application US/08485284A
; Patent No. 5750372
;  GENERAL INFORMATION:
;    APPLICANT:  SAKAI, YASUYOSHI
;    APPLICANT:  TANI, YOSHIKI
;    APPLICANT:  SHIBANO, YUJI
;    APPLICANT:  KONDO, HIROTO
;    APPLICANT:  HATANAKA, HARUYO
```

```
;    TITLE OF INVENTION:  NOVEL VECTOR HAVING PROMOTER THAT IS
;    TITLE OF INVENTION:  INDUCIBLE BY METHANOL AND/OR GLYCEROL
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D. C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/485,284A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 43361/1992
;      FILING DATE:  28-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/025,416
;      FILING DATE:  01-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WHITE JR., PAUL E.
;      REGISTRATION NUMBER:  32,011
;      REFERENCE/DOCKET NUMBER:  217755/FPS38209US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1667 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  genomic DNA
US-08-485-284A-1

  Query Match             4.4%;  Score 43.8;  DB 1;  Length 1667;
  Best Local Similarity   53.9%;  Pred. No. 0.95;
  Matches   90;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy        461 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 520
              |||| ||| |||  ||  |||||| ||| |||  ||  |||||| ||| ||  |  |
Db        624 CTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGGCCCGTATT 565

Qy        521 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAG 580
              ||  ||  ||   |  | ||| |||       |  ||  | |   |  |  |||   |
Db        564 TTGGCTGGCTGGCTGACTGGCTGATTGAGAGATTGGGAGATGAGGCACACGGCTGCAGGG 505

Qy        581 GGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTG 627
              ||||  |  || ||  | ||| ||     || | ||| ||| ||
Db        504 TAGACGCAAGTCCAGCTGCACGGCGTGCATCCAGAGATGCGGGGGTG 458


RESULT 83
US-09-252-991A-6106
; Sequence 6106, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 6106
;   LENGTH: 756
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6106

  Query Match             4.4%;  Score 43.6;  DB 4;  Length 756;
  Best Local Similarity   48.1%;  Pred. No. 0.97;
  Matches  247;  Conservative    0;  Mismatches  259;  Indels    8;  Gaps    4;

Qy         67 GCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCG 126
              | | ||| ||  |||| | ||  |     | |  |     | |   | |     |    |
Db          1 GAGGCGCTGGTCGGGGCGGCGCCGTTGCTGCCCGCGGCAGGCGTTGCCGAAGCAGCGCTG 60

Qy        127 ACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGA 186
                | ||   |||||     ||| |     | |  |      | ||  | ||    |||
Db         61 GAGGCGGCTGCCGGGGCGCTGGGAGTCGCCGCAGGCTTCGCAGCAGGAGCGGTGGCGGGC 120

Qy        187 CGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTT 246
                |  || |   | ||  | | |  |||||   || ||  || ||  ||| ||  |
Db        121 TTGGTTGCGGCAGGTTTCGCGGCTACGGGTTTCGCCGCAGGCTTGGCCGCCGGTTTCGCC 180

Qy        247 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 306
              || |      |||  ||| ||  |||  |||   | ||  || ||     |   |||| |
Db        181 GCAGGCTTCGCAGCCGGCTTGGCTGCAGGCTTCGCGGC-CGCGGATTTCGCCACAGGCTT 239

Qy        307 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG---GACAG 363
              |   || |       | | |||| |    ||| ||| | |||  |||  | |   |
Db        240 GGCCGCCGGTTTCGCCGCAGGCTTCGCAGCTGGCTTGGCTGCGGGCTTCGCGGCTGCGGT 299

Qy        364 TAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 423
              |   || || ||||||  || ||  |    | ||| |   |||  ||| |  |||  |
Db        300 TTTCGCCGCGGGGCTTGGCCGTCGGTTTCGCGACAGGCTTCGCTGCCGGCTTGGCGGCGGG 359

Qy        424 CTT---GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 480
              |||   |  |||    |  | ||  |||||  ||| ||  |   |  |||     ||
Db        360 CTTCGCGGCTGCGGTTTTCGCCGCAGGCTTGGCCGCCGGTTTCGCCACAGGCTTCGCCGC 419

Qy        481 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 540
              | |||||  |||  |||   | |||    |  ||||  ||  |||||   ||  | |
Db        420 CGGCTTGGCTGCGGGCTTCGCGGCTGCGGTTTTCGC-CGCGGGCTTGGCCGCCGGTTTCG 478

Qy        541 CTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 574
              | ||  |||  |  | | |   ||||||  ||||
Db        479 CAGCAGGCTTCGCCGTCGGCTTGGCTGCGGGCTT 512


RESULT 84
US-09-252-991A-6186
; Sequence 6186, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 6186
;   LENGTH: 912
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6186

  Query Match             4.4%;  Score 43.6;  DB 4;  Length 912;
  Best Local Similarity   48.1%;  Pred. No. 0.98;
```

```
  Matches  247;  Conservative    0;  Mismatches  259;  Indels    8;  Gaps    4;

Qy         67 GCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCG 126
              | | ||| ||  |||| | ||  |    | |  |     | |   | |     |    |
Db        195 GAGGCGCTGGTCGGGGCGGCGCCGTTGCTGCCCGCGGCAGGCGTTGCCGAAGCAGCGCTG 254

Qy        127 ACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGA 186
                | ||   |||||     ||| |     | |  |       | ||  | ||    |||
Db        255 GAGGCGGCTGCCGGGGCGCTGGGAGTCGCCGCAGGCTTCGCAGCAGGAGCGGTGGCGGGC 314

Qy        187 CGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTT 246
                |  || |   | ||  | | |  |||||   || ||  || ||  ||| ||  |
Db        315 TTGGTTGCGGCAGGTTTCGCGGCTACGGGTTTCGCCGCAGGCTTGGCCGCCGGTTTCGCC 374

Qy        247 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 306
              || |       |||  ||| || ||  |||  |||   | ||  || ||      |  |||| |
Db        375 GCAGGCTTCGCAGCCGGCTTGGCTGCAGGCTTCGCGGC-CGCGGATTTCGCCACAGGCTT 433

Qy        307 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG---GACAG 363
              |   || |      | | |||| |    ||| ||| | |||  |||  | |   |
Db        434 GGCCGCCGGTTTCGCCGCAGGCTTCGCAGCTGGCTTGGCTGCGGGCTTCGCGGCTGCGGT 493

Qy        364 TAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 423
              |   || || ||||||   || ||  |     | ||| |   |||  ||| | |||  |
Db        494 TTTCGCCGCGGGCTTGGCCGTCGGTTTCGCGACAGGCTTCGCTGCCGGCTTGGCGGCGGG 553

Qy        424 CTT---GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 480
              |||   |  |||    |   | ||  |||||  ||| ||  |     |  |||     ||
Db        554 CTTCGCGGCTGCGGTTTTCGCCGCAGGCTTGGCCGCCGGTTTCGCCACAGGCTTCGCCGC 613

Qy        481 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 540
              | |||||  |||  |||   | |||     |  ||||  ||  |||||   ||  | |
Db        614 CGGCTTGGCTGCGGGCTTCGCGGCTGCGGTTTTCGC-CGCGGGCTTGGCCGCCGGTTTCG 672

Qy        541 CTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 574
              | ||  |||  |  | | |   ||||||  ||||
Db        673 CAGCAGGCTTCGCCGTCGGCTTGGCTGCGGGCTT 706


RESULT 85
US-09-252-991A-6445/c
; Sequence 6445, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 6445
;   LENGTH: 915
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6445

  Query Match             4.4%;  Score 43.6;  DB 4;  Length 915;
  Best Local Similarity   48.1%;  Pred. No. 0.98;
  Matches  247;  Conservative    0;  Mismatches  259;  Indels    8;  Gaps    4;

Qy         67 GCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCG 126
              | | ||| ||  |||| | ||  |    | |  |     | |   | |     |    |
Db        732 GAGGCGCTGGTCGGGGCGGCGCCGTTGCTGCCCGCGGCAGGCGTTGCCGAAGCAGCGCTG 673

Qy        127 ACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGA 186
                | ||   |||||     ||| |     | |  |       | ||  | ||    |||
Db        672 GAGGCGGCTGCCGGGGCGCTGGGAGTCGCCGCAGGCTTCGCAGCAGGAGCGGTGGCGGGC 613

Qy        187 CGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTT 246
                |  || |   | ||  | | |  |||||   || ||  || ||  ||| ||  |
Db        612 TTGGTTGCGGCAGGTTTCGCGGCTACGGGTTTCGCCGCAGGCTTGGCCGCCGGTTTCGCC 553

Qy        247 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 306
              || |       |||  ||| || ||  |||  |||   | ||  || ||      |  |||| |
Db        552 GCAGGCTTCGCAGCCGGCTTGGCTGCAGGCTTCGCGGC-CGCGGATTTCGCCACAGGCTT 494

Qy        307 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG---GACAG 363
              |   || |      | | |||| |    ||| ||| | |||  |||  | |   |
Db        493 GGCCGCCGGTTTCGCCGCAGGCTTCGCAGCTGGCTTGGCTGCGGGCTTCGCGGCTGCGGT 434

Qy        364 TAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 423
              |   || || ||||||   || ||  |     | ||| |   |||  ||| | |||  |
Db        433 TTTCGCCGCGGGCTTGGCCGTCGGTTTCGCGACAGGCTTCGCTGCCGGCTTGGCGGCGGG 374

Qy        424 CTT---GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 480
              |||   |  |||    |   | ||  |||||  ||| ||  |     |  |||     ||
Db        373 CTTCGCGGCTGCGGTTTTCGCCGCAGGCTTGGCCGCCGGTTTCGCCACAGGCTTCGCCGC 314

Qy        481 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 540
              | |||||  |||  |||   | |||     |  ||||  ||  |||||   ||  | |
Db        313 CGGCTTGGCTGCGGGCTTCGCGGCTGCGGTTTTCGC-CGCGGGCTTGGCCGCCGGTTTCG 255

Qy        541 CTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 574
              | ||  |||  |  | | |   ||||||  ||||
Db        254 CAGCAGGCTTCGCCGTCGGCTTGGCTGCGGGCTT 221


RESULT 86
US-09-252-991A-6282/c
; Sequence 6282, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 6282
;   LENGTH: 1200
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6282

  Query Match             4.4%;  Score 43.6;  DB 4;  Length 1200;
  Best Local Similarity   48.1%;  Pred. No. 1;
  Matches  247;  Conservative    0;  Mismatches  259;  Indels    8;  Gaps    4;

Qy         67 GCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCG 126
              | | ||| ||  |||| | ||  |    | |  |     | |   | |     |    |
Db       1196 GAGGCGCTGGTCGGGGCGGCGCCGTTGCTGCCCGCGGCAGGCGTTGCCGAAGCAGCGCTG 1137

Qy        127 ACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGA 186
                | ||   |||||     ||| |     | |  |       | ||  | ||    |||
Db       1136 GAGGCGGCTGCCGGGGCGCTGGGAGTCGCCGCAGGCTTCGCAGCAGGAGCGGTGGCGGGC 1077

Qy        187 CGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTT 246
                |  || |   | ||  | | |  |||||   || ||  || ||  ||| ||  |
Db       1076 TTGGTTGCGGCAGGTTTCGCGGCTACGGGTTTCGCCGCAGGCTTGGCCGCCGGTTTCGCC 1017

Qy        247 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 306
              || |       |||  ||| || ||  |||  |||   | ||  || ||      |  |||| |
Db       1016 GCAGGCTTCGCAGCCGGCTTGGCTGCAGGCTTCGCGGC-CGCGGATTTCGCCACAGGCTT 958

Qy        307 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG---GACAG 363
              |   || |      | | |||| |    ||| ||| | |||  |||  | |   |
```

```
Db        957 GGCCGCCGGTTTCGCCGCAGGCTTCGCAGCTGGCTTGGCTGCGGGCTTCGCGGCTGCGGT 898

Qy        364 TAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 423

Db        897 TTTCGCCGCGGGCTTGGCCGTCGGTTTCGCGACAGGCTTCGCTGCCGGCTTGGCGGCGGG 838

Qy        424 CTT---GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 480

Db        837 CTTCGCGGCTGCGGTTTTCGCCGCAGGCTTGGCCGCCGGTTTCGCCACAGGCTTCGCCGC 778

Qy        481 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 540

Db        777 CGGCTTGGCTGCGGGCTTCGCGGCTGCGGTTTTCGC-CGCGGGCTTGGCCGCCGGTTTCG 719

Qy        541 CTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 574

Db        718 CAGCAGGCTTCGCCGTCGGCTTGGCTGCGGGCTT 685


RESULT 87
US-09-252-991A-6023
; Sequence 6023, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 6023
;   LENGTH: 1806
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6023

  Query Match             4.4%;  Score 43.6;  DB 4;  Length 1806;
  Best Local Similarity   48.1%;  Pred. No. 1;
  Matches  247;  Conservative    0;  Mismatches  259;  Indels    8;  Gaps    4;

Qy         67 GCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCG 126

Db        539 GAGGCGCTGGTCGGGGCGGCGCCGTTGCTGCCCGCGGCAGGCGTTGCCGAAGCAGCGCTG 598

Qy        127 ACCGAGGGCCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGA 186

Db        599 GAGGCGGCTGCCGGGGCGCTGGGAGTCGCCGCAGGCTTCGCAGCAGGAGCGGTGGCGGGC 658

Qy        187 CGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTT 246

Db        659 TTGGTTGCGGCAGGTTTCGCGGCTACGGGTTTCGCCGCAGGCTTGGCCGCCGGTTTCGCC 718

Qy        247 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 306

Db        719 GCAGGCTTCGCAGCCGGCTTGGCTGCAGGCTTCGCGGC-CGCGGATTTCGCCACAGGCTT 777

Qy        307 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG---GACAG 363

Db        778 GGCCGCCGGTTTCGCCGCAGGCTTCGCAGCTGGCTTGGCTGCGGGCTTCGCGGCTGCGGT 837

Qy        364 TAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 423

Db        838 TTTCGCCGCGGGCTTGGCCGTCGGTTTCGCGACAGGCTTCGCTGCCGGCTTGGCGGCGGG 897

Qy        424 CTT---GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 480

Db        898 CTTCGCGGCTGCGGTTTTCGCCGCAGGCTTGGCCGCCGGTTTCGCCACAGGCTTCGCCGC 957

Qy        481 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 540

Db        958 CGGCTTGGCTGCGGGCTTCGCGGCTGCGGTTTTCGC-CGCGGGCTTGGCCGCCGGTTTCG 1016

Qy        541 CTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 574

Db       1017 CAGCAGGCTTCGCCGTCGGCTTGGCTGCGGGCTT 1050


RESULT 88
US-09-776-976-7/c
; Sequence 7, Application US/09776976
; Patent No. 6566332
; GENERAL INFORMATION:
;  APPLICANT: Fruebis, Joachim
;  APPLICANT:  Erickson, Mary Ruth
;  APPLICANT:  Yen, Frances
;  APPLICANT:  Bihain, Bernard
;  TITLE OF INVENTION: OBG3 Globular Head and Uses Thereof for Decreasing Body Mass
;  FILE REFERENCE: 76.US4.REG
;  CURRENT APPLICATION NUMBER: US/09/776,976
;  CURRENT FILING DATE:  2001-02-05
;  PRIOR APPLICATION NUMBER: US 09/758,055
;  PRIOR FILING DATE: 2001-01-10
;  PRIOR APPLICATION NUMBER: US 60/176,228
;  PRIOR FILING DATE: 2000-01-14
;  PRIOR APPLICATION NUMBER: US 60/198,087
;  PRIOR FILING DATE: 2000-04-13
;  PRIOR APPLICATION NUMBER: US 60/299,881
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: Patent.pm
; SEQ ID NO 7
;   LENGTH: 20966
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..4811
;   OTHER INFORMATION: 5' regulatory region
;   NAME/KEY: exon
;   LOCATION: 4812..4851
;   OTHER INFORMATION: exon 1
;   NAME/KEY: exon
;   LOCATION: 15144..15365
;   OTHER INFORMATION: exon 2
;   NAME/KEY: exon
;   LOCATION: 16277..20559
;   OTHER INFORMATION: exon 3
;   NAME/KEY: misc_feature
;   LOCATION: 20560..20966
;   OTHER INFORMATION: 3' regulatory region
;   NAME/KEY: allele
;   LOCATION: 3787
;   OTHER INFORMATION: 9-27-261 : polymorphic base G or C
;   NAME/KEY: allele
;   LOCATION: 11118
;   OTHER INFORMATION: 99-14387-129 : polymorphic base A or C
;   NAME/KEY: allele
;   LOCATION: 15120
;   OTHER INFORMATION: 9-12-48 : polymorphic base C or T
;   NAME/KEY: allele
;   LOCATION: 15196
;   OTHER INFORMATION: 9-12-124 : polymorphic base G or T
;   NAME/KEY: allele
;   LOCATION: 15427
;   OTHER INFORMATION: 9-12-355 : polymorphic base G or T
;   NAME/KEY: allele
;   LOCATION: 15500
;   OTHER INFORMATION: 9-12-428 : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 15863
;   OTHER INFORMATION: 99-14405-105 : polymorphic base A or G
```

```
;   NAME/KEY: allele
;   LOCATION: 17170
;   OTHER INFORMATION: 9-16-189 : polymorphic base deletion of A
;   NAME/KEY: primer_bind
;   LOCATION: 3528..3545
;   OTHER INFORMATION: 9-27.pu
;   NAME/KEY: primer_bind
;   LOCATION: 3928..3946
;   OTHER INFORMATION: 9-27.rp complement
;   NAME/KEY: primer_bind
;   LOCATION: 10990..11008
;   OTHER INFORMATION: 99-14387.pu
;   NAME/KEY: primer_bind
;   LOCATION: 11423..11442
;   OTHER INFORMATION: 99-14387.rp complement
;   NAME/KEY: primer_bind
;   LOCATION: 15073..15092
;   OTHER INFORMATION: 9-12.pu
;   NAME/KEY: primer_bind
;   LOCATION: 15503..15520
;   OTHER INFORMATION: 9-12.rp complement
;   NAME/KEY: primer_bind
;   LOCATION: 15759..15776
;   OTHER INFORMATION: 99-14405.pu
;   NAME/KEY: primer_bind
;   LOCATION: 16191..16211
;   OTHER INFORMATION: 99-14405.rp complement
;   NAME/KEY: primer_bind
;   LOCATION: 16982..17001
;   OTHER INFORMATION: 9-16.pu
;   NAME/KEY: primer_bind
;   LOCATION: 17384..17402
;   OTHER INFORMATION: 9-16.rp complement
;   NAME/KEY: misc_binding
;   LOCATION: 3775..3799
;   OTHER INFORMATION: 9-27-261.probe
;   NAME/KEY: misc_binding
;   LOCATION: 11106..11130
;   OTHER INFORMATION: 99-14387-129.probe
;   NAME/KEY: misc_binding
;   LOCATION: 15108..15132
;   OTHER INFORMATION: 9-12-48.probe
;   NAME/KEY: misc_binding
;   LOCATION: 15184..15208
;   OTHER INFORMATION: 9-12-124.probe
;   NAME/KEY: misc_binding
;   LOCATION: 15415..15439
;   OTHER INFORMATION: 9-12-355.probe
;   NAME/KEY: misc_binding
;   LOCATION: 15488..15512
;   OTHER INFORMATION: 9-12-428.probe
;   NAME/KEY: misc_binding
;   LOCATION: 15851..15875
;   OTHER INFORMATION: 99-14405-105.probe
;   NAME/KEY: misc_binding
;   LOCATION: 17158..17182
;   OTHER INFORMATION: 9-16-189.probe
;   NAME/KEY: primer_bind
;   LOCATION: 3768..3786
;   OTHER INFORMATION: 9-27-261.mis
;   NAME/KEY: primer_bind
;   LOCATION: 3788..3806
;   OTHER INFORMATION: 9-27-261.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 11099..11117
;   OTHER INFORMATION: 99-14387-129.mis
;   NAME/KEY: primer_bind
;   LOCATION: 11119..11137
;   OTHER INFORMATION: 99-14387-129.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 15101..15119
;   OTHER INFORMATION: 9-12-48.mis
;   NAME/KEY: primer_bind
;   LOCATION: 15121..15139
;   OTHER INFORMATION: 9-12-48.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 15177..15195
;   OTHER INFORMATION: 9-12-124.mis
;   NAME/KEY: primer_bind
;   LOCATION: 15197..15215
;   OTHER INFORMATION: 9-12-124.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 15408..15426
;   OTHER INFORMATION: 9-12-355.mis
;   NAME/KEY: primer_bind
;   LOCATION: 15428..15446
;   OTHER INFORMATION: 9-12-355.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 15481..15499
;   OTHER INFORMATION: 9-12-428.mis
;   NAME/KEY: primer_bind
;   LOCATION: 15501..15519
;   OTHER INFORMATION: 9-12-428.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 15844..15862
;   OTHER INFORMATION: 99-14405-105.mis
;   NAME/KEY: primer_bind
;   LOCATION: 15864..15882
;   OTHER INFORMATION: 99-14405-105.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 17151..17169
;   OTHER INFORMATION: 9-16-189.mis
;   NAME/KEY: primer_bind
;   LOCATION: 17171..17189
;   OTHER INFORMATION: 9-16-189.mis complement
US-09-776-976-7

  Query Match             4.4%;  Score 43.6;  DB 4;  Length 20966;
  Best Local Similarity   57.2%;  Pred. No. 1.3;
  Matches   79;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        400 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 459
              | | ||| ||| ||  ||  | |  || ||| ||  ||    | ||| ||| ||   |
Db      10425 CTTCCTTCCTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCTTTC 10366

Qy        460 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 519
              | | ||| ||| ||  ||    | ||| ||| ||  ||    | ||| ||| ||  ||
Db      10365 CTTCCTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCCTTC 10306

Qy        520 CCTGCTTGCTTGCTCGCT 537
              | | |||  || ||| ||
Db      10305 CTTCCTTCTTTTCTCTCT 10288


RESULT 89
US-09-909-547-7/c
; Sequence 7, Application US/09909547
; Patent No. 6579852
; GENERAL INFORMATION:
;  APPLICANT: Fruebis, Joachim
;  APPLICANT:  Erickson, Mary Ruth
;  APPLICANT:  Yen, Frances
;  APPLICANT:  Bihain, Bernard
;  TITLE OF INVENTION: OBG3 Globular Head and Uses Thereof for Decreasing Body Mass
;  FILE REFERENCE: 76.US6.CIP
;  CURRENT APPLICATION NUMBER: US/09/909,547
;  CURRENT FILING DATE:  2001-07-19
;  PRIOR APPLICATION NUMBER: US 09/776,976
;  PRIOR FILING DATE: 2001-02-05
;  PRIOR APPLICATION NUMBER: US 09/758,055
;  PRIOR FILING DATE: 2001-01-10
;  PRIOR APPLICATION NUMBER: US 60/299,881
;  PRIOR FILING DATE: 2000-09-01
;  PRIOR APPLICATION NUMBER: US 60/198,087
;  PRIOR FILING DATE: 2000-04-13
;  PRIOR APPLICATION NUMBER: US 60/176,228
```

; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent.pm
; SEQ ID NO 7
; LENGTH: 20966
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..4811
; OTHER INFORMATION: 5' regulatory region
; NAME/KEY: exon
; LOCATION: 4812..4851
; OTHER INFORMATION: exon 1
; NAME/KEY: exon
; LOCATION: 15144..15365
; OTHER INFORMATION: exon 2
; NAME/KEY: exon
; LOCATION: 16277..20559
; OTHER INFORMATION: exon 3
; NAME/KEY: misc_feature
; LOCATION: 20560..20966
; OTHER INFORMATION: 3' regulatory region
; NAME/KEY: allele
; LOCATION: 3787
; OTHER INFORMATION: 9-27-261 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 11118
; OTHER INFORMATION: 99-14387-129 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 15120
; OTHER INFORMATION: 9-12-48 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 15196
; OTHER INFORMATION: 9-12-124 : polymorphic base G or T
; NAME/KEY: allele
; LOCATION: 15427
; OTHER INFORMATION: 9-12-355 : polymorphic base G or T
; NAME/KEY: allele
; LOCATION: 15500
; OTHER INFORMATION: 9-12-428 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 15863
; OTHER INFORMATION: 99-14405-105 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 17170
; OTHER INFORMATION: 9-16-189 : polymorphic base deletion of A
; NAME/KEY: primer_bind
; LOCATION: 3528..3545
; OTHER INFORMATION: 9-27.pu
; NAME/KEY: primer_bind
; LOCATION: 3928..3946
; OTHER INFORMATION: 9-27.rp complement
; NAME/KEY: primer_bind
; LOCATION: 10990..11008
; OTHER INFORMATION: 99-14387.pu
; NAME/KEY: primer_bind
; LOCATION: 11423..11442
; OTHER INFORMATION: 99-14387.rp complement
; NAME/KEY: primer_bind
; LOCATION: 15073..15092
; OTHER INFORMATION: 9-12.pu
; NAME/KEY: primer_bind
; LOCATION: 15503..15520
; OTHER INFORMATION: 9-12.rp complement
; NAME/KEY: primer_bind
; LOCATION: 15759..15776
; OTHER INFORMATION: 99-14405.pu
; NAME/KEY: primer_bind
; LOCATION: 16191..16211
; OTHER INFORMATION: 99-14405.rp complement
; NAME/KEY: primer_bind
; LOCATION: 16982..17001
; OTHER INFORMATION: 9-16.pu
; NAME/KEY: primer_bind
; LOCATION: 17384..17402
; OTHER INFORMATION: 9-16.rp complement
; NAME/KEY: misc_binding
; LOCATION: 3775..3799
; OTHER INFORMATION: 9-27-261.probe
; NAME/KEY: misc_binding
; LOCATION: 11106..11130
; OTHER INFORMATION: 99-14387-129.probe
; NAME/KEY: misc_binding
; LOCATION: 15108..15132
; OTHER INFORMATION: 9-12-48.probe
; NAME/KEY: misc_binding
; LOCATION: 15184..15208
; OTHER INFORMATION: 9-12-124.probe
; NAME/KEY: misc_binding
; LOCATION: 15415..15439
; OTHER INFORMATION: 9-12-355.probe
; NAME/KEY: misc_binding
; LOCATION: 15488..15512
; OTHER INFORMATION: 9-12-428.probe
; NAME/KEY: misc_binding
; LOCATION: 15851..15875
; OTHER INFORMATION: 99-14405-105.probe
; NAME/KEY: misc_binding
; LOCATION: 17158..17182
; OTHER INFORMATION: 9-16-189.probe
; NAME/KEY: primer_bind
; LOCATION: 3768..3786
; OTHER INFORMATION: 9-27-261.mis
; NAME/KEY: primer_bind
; LOCATION: 3788..3806
; OTHER INFORMATION: 9-27-261.mis complement
; NAME/KEY: primer_bind
; LOCATION: 11099..11117
; OTHER INFORMATION: 99-14387-129.mis
; NAME/KEY: primer_bind
; LOCATION: 11119..11137
; OTHER INFORMATION: 99-14387-129.mis complement
; NAME/KEY: primer_bind
; LOCATION: 15101..15119
; OTHER INFORMATION: 9-12-48.mis
; NAME/KEY: primer_bind
; LOCATION: 15121..15139
; OTHER INFORMATION: 9-12-48.mis complement
; NAME/KEY: primer_bind
; LOCATION: 15177..15195
; OTHER INFORMATION: 9-12-124.mis
; NAME/KEY: primer_bind
; LOCATION: 15197..15215
; OTHER INFORMATION: 9-12-124.mis complement
; NAME/KEY: primer_bind
; LOCATION: 15408..15426
; OTHER INFORMATION: 9-12-355.mis
; NAME/KEY: primer_bind
; LOCATION: 15428..15446
; OTHER INFORMATION: 9-12-355.mis complement
; NAME/KEY: primer_bind
; LOCATION: 15481..15499
; OTHER INFORMATION: 9-12-428.mis
; NAME/KEY: primer_bind
; LOCATION: 15501..15519
; OTHER INFORMATION: 9-12-428.mis complement
; NAME/KEY: primer_bind
; LOCATION: 15844..15862
; OTHER INFORMATION: 99-14405-105.mis
; NAME/KEY: primer_bind
; LOCATION: 15864..15882
; OTHER INFORMATION: 99-14405-105.mis complement
; NAME/KEY: primer_bind
; LOCATION: 17151..17169
; OTHER INFORMATION: 9-16-189.mis

```
;   NAME/KEY: primer_bind
;   LOCATION: 17171..17189
;   OTHER INFORMATION: 9-16-189.mis complement
US-09-909-547-7

  Query Match             4.4%;  Score 43.6;  DB 4;  Length 20966;
  Best Local Similarity   57.2%;  Pred. No. 1.3;
  Matches   79;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        400 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 459
              | | ||| ||| ||  ||  | |  || ||| ||  ||    | ||| ||| ||   |
Db      10425 CTTCCTTCCTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCTTTC 10366

Qy        460 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 519
              | | ||| ||| ||  ||    | ||| ||| ||  ||    | ||| ||| ||  ||
Db      10365 CTTCCTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCCTTC 10306

Qy        520 CCTGCTTGCTTGCTCGCT 537
              | | |||  || ||| ||
Db      10305 CTTCCTTCTTTTCTCTCT 10288


RESULT 90
US-09-569-852B-1/c
; Sequence 1, Application US/09569852B
; Patent No. 6582909
; GENERAL INFORMATION:
;  APPLICANT: Bougueleret, Lydie
;  APPLICANT:  Bihain, Bernard
;  APPLICANT:  Denison, Blake
;  APPLICANT:  Yen-Potin, Frances
;  TITLE OF INVENTION: APM1 Biallelic Markers and Uses Thereof
;  FILE REFERENCE: GEN-T113XC2
;  CURRENT APPLICATION NUMBER: US/09/569,852B
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: PCT/IB99/01858
;  PRIOR FILING DATE: 1999-11-04
;  PRIOR APPLICATION NUMBER: US 09/434,848
;  PRIOR FILING DATE: 1999-11-04
;  PRIOR APPLICATION NUMBER: US 60/119,593
;  PRIOR FILING DATE: 1999-02-10
;  PRIOR APPLICATION NUMBER: US 60/107,113
;  PRIOR FILING DATE: 1998-11-04
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 20966
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(4811)
;   OTHER INFORMATION: 5' regulatory region
;   NAME/KEY: primer_bind
;   LOCATION: (14683)..(14701)
;   OTHER INFORMATION: 17-34-860.mis
;   NAME/KEY: primer_bind
;   LOCATION: (14703)..(14721)
;   OTHER INFORMATION: 17-34-860.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (14738)..(14756)
;   OTHER INFORMATION: 17-34-915.mis
;   NAME/KEY: primer_bind
;   LOCATION: (14758)..(14776)
;   OTHER INFORMATION: 17-34-915.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (14796)..(14814)
;   OTHER INFORMATION: 17-35-71.mis
;   NAME/KEY: primer_bind
;   LOCATION: (14816)..(14834)
;   OTHER INFORMATION: 17-35-71.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (15031)..(15049)
;   OTHER INFORMATION: 17-35-306.mis
;   NAME/KEY: primer_bind
;   LOCATION: (15051)..(15069)
;   OTHER INFORMATION: 17-35-306.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (15101)..(15119)
;   OTHER INFORMATION: 9-12-48.mis
;   NAME/KEY: primer_bind
;   LOCATION: (15121)..(15139)
;   OTHER INFORMATION: 9-12-48.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (15177)..(15195)
;   OTHER INFORMATION: 9-12-124.mis
;   NAME/KEY: primer_bind
;   LOCATION: (15197)..(15215)
;   OTHER INFORMATION: 9-12-124.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (15408)..(15426)
;   OTHER INFORMATION: 9-12-355.mis
;   NAME/KEY: primer_bind
;   LOCATION: (15428)..(15446)
;   OTHER INFORMATION: 9-12-355.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (15481)..(15499)
;   OTHER INFORMATION: 9-12-428.mis
;   NAME/KEY: primer_bind
;   LOCATION: (15501)..(15519)
;   OTHER INFORMATION: 9-12-428.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (15661)..(15679)
;   OTHER INFORMATION: 17-36-47.mis
;   NAME/KEY: primer_bind
;   LOCATION: (15681)..(15699)
;   OTHER INFORMATION: 17-36-47.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (15771)..(15789)
;   OTHER INFORMATION: 17-36-120.mis
;   NAME/KEY: primer_bind
;   LOCATION: (15791)..(15809)
;   OTHER INFORMATION: 17-36-120.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (15844)..(15862)
;   OTHER INFORMATION: 99-14405-105.mis
;   NAME/KEY: primer_bind
;   LOCATION: (15864)..(15882)
;   OTHER INFORMATION: 99-14405-105.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (17151)..(17169)
;   OTHER INFORMATION: 9-16-189.mis
;   NAME/KEY: primer_bind
;   LOCATION: (17171)..(17189)
;   OTHER INFORMATION: 9-16-189.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (17810)..(17828)
;   OTHER INFORMATION: 17-37-629.mis
;   NAME/KEY: primer_bind
;   LOCATION: (17830)..(17848)
;   OTHER INFORMATION: 17-37-629.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (17992)..(18010)
;   OTHER INFORMATION: 17-37-811.mis
;   NAME/KEY: primer_bind
;   LOCATION: (18012)..(18030)
;   OTHER INFORMATION: 17-37-811.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (18470)..(18488)
;   OTHER INFORMATION: 17-38-349.mis
;   NAME/KEY: primer_bind
;   LOCATION: (18490)..(18508)
;   OTHER INFORMATION: 17-38-349.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (926)..(944)
;   OTHER INFORMATION: 17-30-216.mis
```

```
;   NAME/KEY: primer_bind
;   LOCATION: (946)..(964)
;   OTHER INFORMATION: 17-30-216.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (3719)..(3737)
;   OTHER INFORMATION: 9-27-211.mis
;   NAME/KEY: primer_bind
;   LOCATION: (3739)..(3757)
;   OTHER INFORMATION: 9-27-211.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (3754)..(3772)
;   OTHER INFORMATION: 9-27-246.mis
;   NAME/KEY: primer_bind
;   LOCATION: (3774)..(3792)
;   OTHER INFORMATION: 9-27-246.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (3768)..(3786)
;   OTHER INFORMATION: 9-27-261.mis
;   NAME/KEY: primer_bind
;   LOCATION: (3788)..(3806)
;   OTHER INFORMATION: 9-27-261.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (5076)..(5094)
;   OTHER INFORMATION: 17-31-298.mis
;   NAME/KEY: primer_bind
;   LOCATION: (5096)..(5114)
;   OTHER INFORMATION: 17-31-298.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (5191)..(5209)
;   OTHER INFORMATION: 17-31-413.mis
;   NAME/KEY: primer_bind
;   LOCATION: (5211)..(5229)
;   OTHER INFORMATION: 17-31-413.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (5364)..(5385)
;   OTHER INFORMATION: 17-31.rp complement
;   NAME/KEY: primer_bind
;   LOCATION: (10618)..(10636)
;   OTHER INFORMATION: 17-32-24.mis
;   NAME/KEY: primer_bind
;   LOCATION: (10638)..(10656)
;   OTHER INFORMATION: 17-32-24.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (11020)..(11038)
;   OTHER INFORMATION: 99-14387-50.mis
;   NAME/KEY: primer_bind
;   LOCATION: (11040)..(11058)
;   OTHER INFORMATION: 99-14387-50.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (11099)..(11117)
;   OTHER INFORMATION: 99-14387-129.mis
;   NAME/KEY: primer_bind
;   LOCATION: (11119)..(11137)
;   OTHER INFORMATION: 99-14387-129.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (11169)..(11187)
;   OTHER INFORMATION: 99-14387-199.mis
;   NAME/KEY: primer_bind
;   LOCATION: (11189)..(11207)
;   OTHER INFORMATION: 99-14387-199.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: (13954)..(13972)
;   OTHER INFORMATION: 17-33-TGAGACT.mis
;   NAME/KEY: primer_bind
;   LOCATION: (13974)..(13992)
;   OTHER INFORMATION: 17-33-TGAGACT.mis complement
;   NAME/KEY: exon
;   LOCATION: (4812)..(4851)
;   OTHER INFORMATION:
;   NAME/KEY: exon
;   LOCATION: (15144)..(15365)
;   OTHER INFORMATION:
;   NAME/KEY: exon
;   LOCATION: (16277)..(20559)
;   OTHER INFORMATION:
;   NAME/KEY: misc_feature
;   LOCATION: (20560)..(20966)
;   OTHER INFORMATION: 3' regulatory region

  Query Match             4.4%;  Score 43.6;  DB 4;  Length 20966;
  Best Local Similarity   57.2%;  Pred. No. 1.3;
  Matches   79;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        400 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 459
              | | ||| ||| ||  ||  | |  || ||| ||  ||    | ||| ||| ||   |
Db      10425 CTTCCTTCCTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCTTTC 10366

Qy        460 CCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 519
              | | ||| ||| ||  ||    | ||| ||| ||  ||    | ||| ||| ||  ||
Db      10365 CTTCCTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCCTTCTTTCCTTCCTTCCTTCCTTC 10306

Qy        520 CCTGCTTGCTTGCTCGCT 537
              | | |||  || ||| ||
Db      10305 CTTCCTTCTTTTCTCTCT 10288


RESULT 91
US-09-179-558-64/c
; Sequence 64, Application US/09179558
; Patent No. 6180612
;  GENERAL INFORMATION:
;    APPLICANT:  Hockensmith, Joel W.
;    APPLICANT:  Muthuswami, Rohini
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR
;    TITLE OF INVENTION:  TARGETING DNA METABOLIC PROCESSES USING
;    TITLE OF INVENTION:  AMINOGLYCOSIDE DERIVATIVES
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/179,558
;      FILING DATE:  27-OCT-1998
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.S. 09/060,470
;      FILING DATE:  15-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.S. 60/063,898
;      FILING DATE:  31-OCT-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  9426-005-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)7909090
;      TELEFAX:  (212)8699741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  64:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  892 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Other
US-09-179-558-64
```

```
  Query Match             4.3%;  Score 43.2;  DB 3;  Length 892;
  Best Local Similarity   56.2%;  Pred. No. 1.2;
  Matches   81;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy         64 TCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGC 123
              | || | | |||||  ||| |     |   ||  || | | | |||  | || | |||||
Db        146 TTCGGGTCCCCGGGTTGGGAGAACCCGAGTGGTCCGGCCCCGGGGGTGCGGGTGGGGGGC 87

Qy        124 GCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGG 183
              | |   | |    | ||| |   |  ||   ||||||| | | |||| | |||||| |
Db         86 GGGGGGGGGAAGTCGGGGGGGTGGAAGGGAGGGGAAAAAGGGCGGGGGGGGGGGGAAAAA 27

Qy        184 GGACGGCGTGTGCGTGCTTGTGAG 207
              || ||| | | | | |   | |||
Db         26 GGCCGGGGGGGGGGGGGGGGGAG 3


RESULT 92
US-09-722-825-64/c
; Sequence 64, Application US/09722825
; Patent No. 6531306
;   GENERAL INFORMATION:
;        APPLICANT: Hockensmith, Joel W.
;                   Muthuswami, Rohini
;        TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
;                            TARGETING DNA METABOLIC PROCESSES USING
;                            AMINOGLYCOSIDE DERIVATIVES
;        NUMBER OF SEQUENCES: 66
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: PENNIE & EDMONDS LLP
;             STREET: 1155 Avenue of the Americas
;             CITY: New York
;             STATE: NY
;             COUNTRY: USA
;             ZIP: 10036-2711
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/722,825
;             FILING DATE: 28-No. 6531306-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/179,558
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: U.S. 60/063,898
;             FILING DATE: 31-OCT-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Coruzzi, Laura A
;             REGISTRATION NUMBER: 30,742
;             REFERENCE/DOCKET NUMBER: 9426-005-999
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (212)7909090
;             TELEFAX: (212)8699741
;             TELEX: 66141 PENNIE
;   INFORMATION FOR SEQ ID NO: 64:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 892 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: Other
;        SEQUENCE DESCRIPTION: SEQ ID NO: 64:
US-09-722-825-64

  Query Match             4.3%;  Score 43.2;  DB 4;  Length 892;
  Best Local Similarity   56.2%;  Pred. No. 1.2;
  Matches   81;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy         64 TCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGC 123
              | || | | |||||  ||| |     |   ||  || | | | |||  | || | |||||

Db        146 TTCGGGTCCCCGGGTTGGGAGAACCCGAGTGGTCCGGCCCCGGGGGTGCGGGTGGGGGGC 87

Qy        124 GCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGG 183
              | |   | |    | ||| |   |  ||   ||||||| | | |||| | |||||| |
Db         86 GGGGGGGGGAAGTCGGGGGGGTGGAAGGGAGGGGAAAAAGGGCGGGGGGGGGGGGAAAAA 27

Qy        184 GGACGGCGTGTGCGTGCTTGTGAG 207
              || ||| | | | | |   | |||
Db         26 GGCCGGGGGGGGGGGGGGGGGAG 3


RESULT 93
US-09-722-487-64/c
; Sequence 64, Application US/09722487
; Patent No. 6537791
;   GENERAL INFORMATION:
;        APPLICANT: Hockensmith, Joel W.
;                   Muthuswami, Rohini
;        TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
;                            TARGETING DNA METABOLIC PROCESSES USING
;                            AMINOGLYCOSIDE DERIVATIVES
;        NUMBER OF SEQUENCES: 66
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: PENNIE & EDMONDS LLP
;             STREET: 1155 Avenue of the Americas
;             CITY: New York
;             STATE: NY
;             COUNTRY: USA
;             ZIP: 10036-2711
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/722,487
;             FILING DATE: 28-No. 6537791-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/179,558
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: U.S. 60/063,898
;             FILING DATE: 31-OCT-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Coruzzi, Laura A
;             REGISTRATION NUMBER: 30,742
;             REFERENCE/DOCKET NUMBER: 9426-005-999
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (212)7909090
;             TELEFAX: (212)8699741
;             TELEX: 66141 PENNIE
;   INFORMATION FOR SEQ ID NO: 64:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 892 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: Other
;        SEQUENCE DESCRIPTION: SEQ ID NO: 64:
US-09-722-487-64

  Query Match             4.3%;  Score 43.2;  DB 4;  Length 892;
  Best Local Similarity   56.2%;  Pred. No. 1.2;
  Matches   81;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy         64 TCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGC 123
              | || | | |||||  ||| |     |   ||  || | | | |||  | || | |||||

Db        146 TTCGGGTCCCCGGGTTGGGAGAACCCGAGTGGTCCGGCCCCGGGGGTGCGGGTGGGGGGC 87

Qy        124 GCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGG 183
              | |   | |    | ||| |   |  ||   ||||||| | | |||| | |||||| |
Db         86 GGGGGGGGGAAGTCGGGGGGGTGGAAGGGAGGGGAAAAAGGGCGGGGGGGGGGGGAAAAA 27
```

```
Qy        184 GGACGGCGTGTGCGTGCTTGTGAG 207
              || ||| | | | | |   | |||
Db         26 GGCCGGGGGGGGGGGGGGGGGAG 3


RESULT 94
US-09-722-708-64/c
; Sequence 64, Application US/09722708
; Patent No. 6573060
;   GENERAL INFORMATION:
;        APPLICANT: Hockensmith, Joel W.
;                   Muthuswami, Rohini
;        TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
;                            TARGETING DNA METABOLIC PROCESSES USING
;                            AMINOGLYCOSIDE DERIVATIVES
;        NUMBER OF SEQUENCES: 66
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: PENNIE & EDMONDS LLP
;             STREET: 1155 Avenue of the Americas
;             CITY: New York
;             STATE: NY
;             COUNTRY: USA
;             ZIP: 10036-2711
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/722,708
;             FILING DATE: 28-No. 6573060-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/179,558
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: U.S. 60/063,898
;             FILING DATE: 31-OCT-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Coruzzi, Laura A
;             REGISTRATION NUMBER: 30,742
;             REFERENCE/DOCKET NUMBER: 9426-005-999
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (212)7909090
;             TELEFAX: (212)8699741
;             TELEX: 66141 PENNIE
;   INFORMATION FOR SEQ ID NO: 64:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 892 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: Other
;        SEQUENCE DESCRIPTION: SEQ ID NO: 64:
US-09-722-708-64

  Query Match             4.3%;  Score 43.2;  DB 4;  Length 892;
  Best Local Similarity   56.2%;  Pred. No. 1.2;
  Matches   81;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy         64 TCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGC 123
              | || | | ||||  ||| |     |   ||  || | | | |||  | || | |||||
Db        146 TTCGGGTCCCCGGGTTGGGAGAACCCGAGTGGTCCGGCCCCGGGGGTGCGGGTGGGGGGC 87

Qy        124 GCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGG 183
              | |   | |    | ||| |   |  ||   ||||||| | | |||| | |||||| |
Db         86 GGGGGGGGGAAGTCGGGGGGGTGGAAGGGAGGGGAAAAAGGGCGGGGGGGGGGGGAAAAA 27

Qy        184 GGACGGCGTGTGCGTGCTTGTGAG 207
              || ||| | | | | |   | |||
Db         26 GGCCGGGGGGGGGGGGGGGGGAG 3


RESULT 95
US-09-179-558-62/c
; Sequence 62, Application US/09179558
; Patent No. 6180612
;  GENERAL INFORMATION:
;    APPLICANT:  Hockensmith, Joel W.
;    APPLICANT:  Muthuswami, Rohini
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR
;    TITLE OF INVENTION:  TARGETING DNA METABOLIC PROCESSES USING
;    TITLE OF INVENTION:  AMINOGLYCOSIDE DERIVATIVES
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/179,558
;      FILING DATE:  27-OCT-1998
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.S. 09/060,470
;      FILING DATE:  15-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.S. 60/063,898
;      FILING DATE:  31-OCT-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  9426-005-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)7909090
;      TELEFAX:  (212)8699741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  62:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  936 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Other
US-09-179-558-62

  Query Match             4.3%;  Score 43.2;  DB 3;  Length 936;
  Best Local Similarity   56.2%;  Pred. No. 1.2;
  Matches   81;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy         64 TCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGC 123
              | || | | ||||  ||| |     |   ||  || | | | |||  | || | |||||
Db        190 TTCGGGTCCCCGGGTTGGGAGAACCCGAGTGGTCCGGCCCCGGGGGTGCGGGTGGGGGGC 131

Qy        124 GCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGG 183
              | |   | |    | ||| |   |  ||   ||||||| | | |||| | |||||| |
Db        130 GGGGGGGGGAAGTCGGGGGGGTGGAAGGGAGGGGAAAAAGGGCGGGGGGGGGGGGAAAAA 71

Qy        184 GGACGGCGTGTGCGTGCTTGTGAG 207
              || ||| | | | | |   | |||
Db         70 GGCCGGGGGGGGGGGGGGGGGAG 47


RESULT 96
US-09-722-825-62/c
; Sequence 62, Application US/09722825
; Patent No. 6531306
;   GENERAL INFORMATION:
;        APPLICANT: Hockensmith, Joel W.
;                   Muthuswami, Rohini
```

```
;       TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
;                           TARGETING DNA METABOLIC PROCESSES USING
;                           AMINOGLYCOSIDE DERIVATIVES
;       NUMBER OF SEQUENCES: 66
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: PENNIE & EDMONDS LLP
;            STREET: 1155 Avenue of the Americas
;            CITY: New York
;            STATE: NY
;            COUNTRY: USA
;            ZIP: 10036-2711
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: DOS
;            SOFTWARE: FastSEQ Version 2.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/722,825
;            FILING DATE: 28-No. 6531306-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/179,558
;            FILING DATE: <Unknown>
;            APPLICATION NUMBER: U.S. 60/063,898
;            FILING DATE: 31-OCT-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Coruzzi, Laura A
;            REGISTRATION NUMBER: 30,742
;            REFERENCE/DOCKET NUMBER: 9426-005-999
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (212)7909090
;            TELEFAX: (212)8699741
;            TELEX: 66141 PENNIE
;  INFORMATION FOR SEQ ID NO: 62:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 936 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       MOLECULE TYPE: Other
;       SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-09-722-825-62

  Query Match             4.3%;  Score 43.2;  DB 4;  Length 936;
  Best Local Similarity   56.2%;  Pred. No. 1.2;
  Matches   81;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy         64 TCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGC 123
              | || | | |||||  ||| |      |   ||  || | | | |||  | || | |||||
Db        190 TTCGGGTCCCCGGGTTGGGAGAACCCGAGTGGTCCGGCCCCGGGGGTGCGGGTGGGGGGC 131

Qy        124 GCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGG 183
              | |   | |    | ||| |   |  ||   ||||||| | | |||| | ||||||
Db        130 GGGGGGGGGAAGTCGGGGGGGTGGAAGGGAGGGGAAAAAGGGCGGGGGGGGGGGGAAAAA 71

Qy        184 GGACGGCGTGTGCGTGCTTGTGAG 207
              || ||| | | | | |   | |||
Db         70 GGCCGGGGGGGGGGGGGGGGGAG 47


RESULT 97
US-09-722-487-62/c
; Sequence 62, Application US/09722487
; Patent No. 6537791
;   GENERAL INFORMATION:
;       APPLICANT: Hockensmith, Joel W.
;                  Muthuswami, Rohini
;       TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
;                           TARGETING DNA METABOLIC PROCESSES USING
;                           AMINOGLYCOSIDE DERIVATIVES
;       NUMBER OF SEQUENCES: 66
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: PENNIE & EDMONDS LLP
;            STREET: 1155 Avenue of the Americas
;            CITY: New York
;            STATE: NY
;            COUNTRY: USA
;            ZIP: 10036-2711
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: DOS
;            SOFTWARE: FastSEQ Version 2.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/722,487
;            FILING DATE: 28-No. 6537791-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/179,558
;            FILING DATE: <Unknown>
;            APPLICATION NUMBER: U.S. 60/063,898
;            FILING DATE: 31-OCT-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Coruzzi, Laura A
;            REGISTRATION NUMBER: 30,742
;            REFERENCE/DOCKET NUMBER: 9426-005-999
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (212)7909090
;            TELEFAX: (212)8699741
;            TELEX: 66141 PENNIE
;  INFORMATION FOR SEQ ID NO: 62:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 936 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       MOLECULE TYPE: Other
;       SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-09-722-487-62

  Query Match             4.3%;  Score 43.2;  DB 4;  Length 936;
  Best Local Similarity   56.2%;  Pred. No. 1.2;
  Matches   81;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy         64 TCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGC 123
              | || | | |||||  ||| |      |   ||  || | | | |||  | || | |||||
Db        190 TTCGGGTCCCCGGGTTGGGAGAACCCGAGTGGTCCGGCCCCGGGGGTGCGGGTGGGGGGC 131

Qy        124 GCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGG 183
              | |   | |    | ||| |   |  ||   ||||||| | | |||| | ||||||
Db        130 GGGGGGGGGAAGTCGGGGGGGTGGAAGGGAGGGGAAAAAGGGCGGGGGGGGGGGGAAAAA 71

Qy        184 GGACGGCGTGTGCGTGCTTGTGAG 207
              || ||| | | | | |   | |||
Db         70 GGCCGGGGGGGGGGGGGGGGGAG 47


RESULT 98
US-09-722-708-62/c
; Sequence 62, Application US/09722708
; Patent No. 6573060
;   GENERAL INFORMATION:
;       APPLICANT: Hockensmith, Joel W.
;                  Muthuswami, Rohini
;       TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
;                           TARGETING DNA METABOLIC PROCESSES USING
;                           AMINOGLYCOSIDE DERIVATIVES
;       NUMBER OF SEQUENCES: 66
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: PENNIE & EDMONDS LLP
;            STREET: 1155 Avenue of the Americas
;            CITY: New York
;            STATE: NY
;            COUNTRY: USA
;            ZIP: 10036-2711
;       COMPUTER READABLE FORM:
```

```
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ Version 2.0
;       CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/722,708
;             FILING DATE: 28-No. 6573060-2000
;             CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/179,558
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: U.S. 60/063,898
;             FILING DATE: 31-OCT-1997
;       ATTORNEY/AGENT INFORMATION:
;             NAME: Coruzzi, Laura A
;             REGISTRATION NUMBER: 30,742
;             REFERENCE/DOCKET NUMBER: 9426-005-999
;       TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (212)7909090
;             TELEFAX: (212)8699741
;             TELEX: 66141 PENNIE
;   INFORMATION FOR SEQ ID NO: 62:
;       SEQUENCE CHARACTERISTICS:
;             LENGTH: 936 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;       MOLECULE TYPE: Other
;       SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-09-722-708-62

  Query Match             4.3%;  Score 43.2;  DB 4;  Length 936;
  Best Local Similarity   56.2%;  Pred. No. 1.2;
  Matches   81;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy         64 TCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGC 123
              | || | | ||||| ||| |     |   ||  || | | | ||| | || | |||||
Db        190 TTCGGGTCCCCGGGTTGGGAGAACCCGAGTGGTCCGGCCCCGGGGGTGCGGGTGGGGGGC 131

Qy        124 GCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGG 183
              | |   | |    | ||| |   |  ||   ||||||| | | |||| | |||||| |
Db        130 GGGGGGGGGAAGTCGGGGGGGTGGAAGGGAGGGGAAAAAGGGCGGGGGGGGGGGGAAAAA 71

Qy        184 GGACGGCGTGTGCGTGCTTGTGAG 207
              || ||| | | | | |   | |||
Db         70 GGCCGGGGGGGGGGGGGGGGGAG 47


RESULT 99
US-08-765-907A-1
; Sequence 1, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
;  APPLICANT: BLANC, Veronique
;  APPLICANT:  THIBAUT, Denis
;  APPLICANT:  BAMAS-JACQUES, Nathalie
;  APPLICANT:  BLANCHE, Francis
;  APPLICANT:  COUZET, Joel
;  APPLICANT:  BARRIERE, Jean-Claude
;  APPLICANT:  DEBUSSCHE, Laurent
;  APPLICANT:  FAMECHON, Alain
;  APPLICANT:  PARIS, Jean-Marc
;  APPLICANT:  DUTRUC-ROSSET, Gilles
;  TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
;  TITLE OF INVENTION:  Mutasynthesis
;  FILE REFERENCE: Streptogramin genes
;  CURRENT APPLICATION NUMBER: US/08/765,907A
;  CURRENT FILING DATE:  1997-03-20
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2888
;   TYPE: DNA
;   ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-1

  Query Match             4.3%;  Score 43.2;  DB 4;  Length 2888;
  Best Local Similarity   51.2%;  Pred. No. 1.3;
  Matches  130;  Conservative    0;  Mismatches  118;  Indels    6;  Gaps    1;

Qy          6 CCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCTC 65
              ||||||   | |  |   ||| | ||      |   | || || |  |   || ||
Db        768 CCCGGCGACGGGTCCGCTCGCGGACCGGGTGACGGACCCGGCGGCGGGGCCGGCGGCGGG 827

Qy         66 CGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGC 125
              |  |     ||| ||||    ||| || | | | ||   |   ||| |||||  |||| |
Db        828 CCGGGACGTGGGCCGGGACGTGGGCCCGGCGTCCCCGGCGACCGGCACGGCGGCGGGCCC 887

Qy        126 GACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGG 185
              |  || ||||||| | | |  ||| |     |       | ||  ||  | | |||| | ||||
Db        888 GGACGTGGGCCCGGCGTGCCCGGCGACCGG------CACGGTGGCGGGGCGGGGCGGGGG 941

Qy        186 ACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCT 245
              ||||   ||||  |   ||||    ||| | |     | | || ||||   | | ||||
Db        942 ACGGTCAGTGCAGGGCGGTGAACATCCGCGCGCACAGCCGTTCCAGCTCCGCGCCGTGCT 1001

Qy        246 TGCTGAGGGGACAG 259
              ||  ||   || |
Db       1002 CGCCCAGCACACCG 1015


RESULT 100
US-08-690-473-1
; Sequence 1, Application US/08690473
; Patent No. 5876923
;  GENERAL INFORMATION:
;    APPLICANT:  Leopardi, Rosario
;    APPLICANT:  Roizman, Bernard
;    TITLE OF INVENTION:  HERPES SIMPLEX VIRUS ICP4 AS AN
;    TITLE OF INVENTION:  INHIBITOR OF APOPTOSIS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/690,473
;      FILING DATE:  26-JUL-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Highlander, Steven L.
;      REGISTRATION NUMBER:  37,642
;      REFERENCE/DOCKET NUMBER:  ARCD:239
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  512/418-3000
;      TELEFAX:  512/474-7577
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4257 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-690-473-1

  Query Match             4.3%;  Score 43.2;  DB 2;  Length 4257;
  Best Local Similarity   48.8%;  Pred. No. 1.4;
  Matches  117;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;
```

```
Qy          1 CGCCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTG 60
              |||||||| |    |     |  |  || || || | ||| |  |  ||  |||
Db       1932 CGCCGCCCTGGGGCGGCTGTCCGCCGCGCCCGCCTCCCCCGCGGGGGGCGACGACCCCGA 1991

Qy         61 GTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGG 120
                | | |||| |   || |  | |   | | | ||||||| || |  |||| | ||||
Db       1992 CGCCGCCCGCCACGCCGACGCCGACGACGACGCCGGGCGCCGCGCCCAGGCCGGCCGCGT 2051

Qy        121 GGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGAC 180
              |||      | |     | |     || ||||||  | |  |||  |  ||| |
Db       2052 GGCCGTCGAGTGCCTGGCCGCCTGCCGCGGGATCCTGGAGGCGCTGGCCGAGGGCTTCGA 2111

Qy        181 AGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGC 240
               || |||   | |  ||| |  | |     | |||  ||||   |   |  | ||| | |
Db       2112 CGGCGACCTGGCGGCCGTCCCGGGGCTGGCCGGGGCCCGGCCCGCCAGCCCCCCGCGGCC 2171
```

Search completed: August 22, 2003, 00:42:31
Job time : 112.038 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 21, 2003, 21:15:12 ; Search time 246.623 Seconds
(without alignments)
9124.900 Million cell updates/sec

Title: US-09-831-000-1_COPY_132719_133719
Perfect score: 1001
Sequence: 1 cgccgcccggcatggcggtc.....cgggcacccggcccagatct 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1517243 seqs, 1124081882 residues

Total number of hits satisfying chosen parameters: 3034486

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	193.4	19.3	576	9	US-09-864-761-26582	Sequence 26582, A
	2	174.6	17.4	555	9	US-09-864-761-9605	Sequence 9605, Ap
	3	111.4	11.1	1360	12	US-10-017-161-1345	Sequence 1345, Ap
	4	94.6	9.5	1672	12	US-10-017-161-2117	Sequence 2117, Ap
c	5	82.6	8.3	798	13	US-10-027-632-164203	Sequence 164203,
c	6	82.6	8.3	798	13	US-10-027-632-164204	Sequence 164204,
	7	81.4	8.1	14750	12	US-10-017-161-1915	Sequence 1915, Ap
	8	77.4	7.7	1830121	14	US-10-329-960-1	Sequence 1, Appli
	9	77	7.7	42999	10	US-09-799-462A-17	Sequence 17, Appl
	10	77	7.7	42999	11	US-09-836-911A-17	Sequence 17, Appl
	11	77	7.7	42999	13	US-10-125-767-17	Sequence 17, Appl
	12	77	7.7	42999	14	US-10-151-081-17	Sequence 17, Appl
	13	77	7.7	42999	14	US-10-287-313-17	Sequence 17, Appl
	14	77	7.7	42999	14	US-10-219-694-17	Sequence 17, Appl
c	15	74.8	7.5	428	9	US-09-728-446-369	Sequence 369, App
c	16	74.8	7.5	436	9	US-09-728-446-354	Sequence 354, App
c	17	73.8	7.4	1006	13	US-10-027-632-273001	Sequence 273001,
c	18	73.4	7.3	572	13	US-10-027-632-272985	Sequence 272985,
c	19	73.2	7.3	388	9	US-09-728-446-295	Sequence 295, App
c	20	73.2	7.3	15720	9	US-09-922-217-1058	Sequence 1058, Ap
c	21	73.2	7.3	15720	10	US-09-833-263-1058	Sequence 1058, Ap
c	22	73.2	7.3	15720	13	US-10-025-380-1058	Sequence 1058, Ap
	23	73	7.3	1297	12	US-10-017-161-2193	Sequence 2193, Ap
c	24	72.2	7.2	409	9	US-09-728-446-362	Sequence 362, App
	25	71.6	7.2	3213	13	US-10-027-632-113494	Sequence 113494,
c	26	71.4	7.1	392	9	US-09-728-446-736	Sequence 736, App
c	27	70.8	7.1	379	9	US-09-728-446-246	Sequence 246, App
c	28	70.8	7.1	427	9	US-09-728-446-250	Sequence 250, App
c	29	70.6	7.1	410	9	US-09-728-446-325	Sequence 325, App
	30	69.4	6.9	368	11	US-09-764-891-8904	Sequence 8904, Ap
	31	69.4	6.9	368	11	US-09-764-891-10047	Sequence 10047, A
	32	67	6.7	397	9	US-09-728-446-465	Sequence 465, App
c	33	65.2	6.5	368	13	US-10-040-739-1102	Sequence 1102, Ap
c	34	64.2	6.4	1771	9	US-09-874-923-7	Sequence 7, Appli
c	35	64.2	6.4	1771	10	US-09-991-496-7	Sequence 7, Appli
c	36	63.8	6.4	681	13	US-10-027-632-141458	Sequence 141458,
	37	63.6	6.4	13830	12	US-10-017-161-1581	Sequence 1581, Ap
	38	62.8	6.3	31718	11	US-09-764-872-812	Sequence 812, App
	39	62.8	6.3	31718	11	US-09-764-872-813	Sequence 813, App
	40	62.8	6.3	31718	11	US-09-764-891-9103	Sequence 9103, Ap
	41	62.8	6.3	31718	11	US-09-764-891-9104	Sequence 9104, Ap
c	42	62.6	6.3	681	13	US-10-027-632-141457	Sequence 141457,
c	43	62.6	6.3	681	13	US-10-027-632-141460	Sequence 141460,
	44	62.6	6.3	726	13	US-10-027-632-15974	Sequence 15974, A
c	45	62.2	6.2	681	13	US-10-027-632-141459	Sequence 141459,
c	46	62.2	6.2	2940917	13	US-10-027-632-174763	Sequence 174763,
	47	61.4	6.1	863	13	US-10-027-632-161036	Sequence 161036,
	48	61.4	6.1	863	13	US-10-027-632-161037	Sequence 161037,
	49	61.4	6.1	863	13	US-10-027-632-161038	Sequence 161038,
	50	61	6.1	223	10	US-09-764-877-3185	Sequence 3185, Ap
	51	61	6.1	495	13	US-10-027-632-273402	Sequence 273402,
	52	60.6	6.1	6109	12	US-10-311-455-300	Sequence 300, App
	53	60.2	6.0	588	13	US-10-027-632-143381	Sequence 143381,
	54	60.2	6.0	13240	12	US-10-017-161-1673	Sequence 1673, Ap
	55	59.4	5.9	223	10	US-09-764-877-3186	Sequence 3186, Ap
c	56	59.2	5.9	647	13	US-10-027-632-281951	Sequence 281951,
	57	59	5.9	588	13	US-10-027-632-143383	Sequence 143383,
	58	59	5.9	110079	14	US-10-175-523-96	Sequence 96, Appl
c	59	58.6	5.9	393	9	US-09-728-446-978	Sequence 978, App
	60	58.6	5.9	588	13	US-10-027-632-143382	Sequence 143382,
	61	58.2	5.8	12733	14	US-10-032-393-47	Sequence 47, Appl
	62	58.2	5.8	12739	14	US-10-032-393-8	Sequence 8, Appli
	63	58	5.8	3163	12	US-10-017-161-1857	Sequence 1857, Ap
	64	58	5.8	90650	14	US-10-175-523-80	Sequence 80, Appl
c	65	57.2	5.7	6171	12	US-10-311-455-762	Sequence 762, App
	66	57	5.7	300000	14	US-10-262-552-33	Sequence 33, Appl
c	67	56.8	5.7	3595	10	US-09-089-818B-6	Sequence 6, Appli
c	68	56.6	5.7	412	11	US-09-803-719-1534	Sequence 1534, Ap
c	69	56	5.6	573	9	US-09-864-761-9961	Sequence 9961, Ap
c	70	56	5.6	573	9	US-09-864-761-26463	Sequence 26463, A
c	71	55.8	5.6	159095	14	US-10-017-128-3	Sequence 3, Appli
c	72	55.6	5.6	354	9	US-09-728-446-435	Sequence 435, App
	73	55.6	5.6	489	13	US-10-027-632-170164	Sequence 170164,
	74	55.6	5.6	489	13	US-10-027-632-170165	Sequence 170165,
	75	55.6	5.6	978	13	US-10-027-632-170163	Sequence 170163,
	76	55.2	5.5	631	13	US-10-027-632-174148	Sequence 174148,
	77	55.2	5.5	631	13	US-10-027-632-174149	Sequence 174149,
c	78	54.8	5.5	439	13	US-10-027-632-68339	Sequence 68339, A
c	79	54.8	5.5	439	13	US-10-027-632-68340	Sequence 68340, A
c	80	54.8	5.5	439	13	US-10-027-632-295071	Sequence 295071,
c	81	54.8	5.5	439	13	US-10-027-632-295072	Sequence 295072,
c	82	54.8	5.5	549	13	US-10-027-632-37032	Sequence 37032, A
c	83	54.8	5.5	549	13	US-10-027-632-37033	Sequence 37033, A
c	84	54.6	5.5	707	13	US-10-027-632-141461	Sequence 141461,
	85	54.6	5.5	15203	12	US-10-017-161-1971	Sequence 1971, Ap
	86	54.4	5.4	1117	12	US-10-017-161-1403	Sequence 1403, Ap
c	87	53.8	5.4	345	9	US-09-728-446-190	Sequence 190, App
c	88	53.8	5.4	956	13	US-10-027-632-31508	Sequence 31508, A
c	89	53.8	5.4	2201	12	US-09-791-254-4	Sequence 4, Appli

c	90	53.8	5.4	4065	12	US-09-791-254-1 Sequence 1, Appli
	91	53.8	5.4	170834	9	US-09-835-232-7 Sequence 7, Appli
c	92	52.4	5.2	502	13	US-10-027-632-270367 Sequence 270367,
c	93	52.4	5.2	502	13	US-10-027-632-270368 Sequence 270368,
	94	52.4	5.2	553	13	US-10-027-632-183437 Sequence 183437,
	95	52.4	5.2	3133	12	US-10-017-161-1483 Sequence 1483, Ap
	96	52.2	5.2	1297	12	US-10-017-161-2377 Sequence 2377, Ap
	97	52	5.2	7597	12	US-10-311-455-985 Sequence 985, App
	98	51.8	5.2	4403	10	US-09-880-107-2402 Sequence 2402, Ap
c	99	51.6	5.2	1360	12	US-10-017-161-1345 Sequence 1345, Ap
c	100	51.2	5.1	582	13	US-10-027-632-269946 Sequence 269946,
c	101	51	5.1	369	10	US-09-867-701-5268 Sequence 5268, Ap
	102	50.8	5.1	802	14	US-10-184-644-312 Sequence 312, App
	103	50.8	5.1	802	14	US-10-184-634-312 Sequence 312, App
c	104	50.8	5.1	12007	12	US-10-311-455-689 Sequence 689, App
c	105	50.6	5.1	589	13	US-10-027-632-86169 Sequence 86169, A
c	106	50.6	5.1	636	13	US-10-027-632-308921 Sequence 308921,
c	107	50.6	5.1	636	13	US-10-027-632-308922 Sequence 308922,
c	108	50.6	5.1	30000	14	US-10-007-010-10 Sequence 10, Appl
	109	50.4	5.0	2681	14	US-10-232-563-1 Sequence 1, Appli
c	110	50.4	5.0	8367	12	US-10-017-161-807 Sequence 807, App
c	111	50.4	5.0	8459	9	US-09-817-913-8 Sequence 8, Appli
c	112	50.4	5.0	8459	10	US-09-817-538-8 Sequence 8, Appli
c	113	50.4	5.0	8459	11	US-09-563-728A-31 Sequence 31, Appl
c	114	50.2	5.0	2000	10	US-09-938-842A-5032 Sequence 5032, Ap
c	115	50.2	5.0	155074	13	US-10-026-188-6 Sequence 6, Appli
	116	50	5.0	766	13	US-10-027-632-30729 Sequence 30729, A
	117	50	5.0	766	13	US-10-027-632-30730 Sequence 30730, A
	118	50	5.0	22400	14	US-10-143-266-1 Sequence 1, Appli
	119	50	5.0	1223197	13	US-10-027-632-179264 Sequence 179264,
	120	49.6	5.0	26555	10	US-09-860-670-161 Sequence 161, App
	121	49.6	5.0	136328	12	US-10-101-510-127 Sequence 127, App
c	122	49.4	4.9	519	9	US-09-864-761-12571 Sequence 12571, A
	123	49.4	4.9	3593	13	US-10-002-775-10 Sequence 10, Appl
	124	49.4	4.9	3593	14	US-10-115-615-22 Sequence 22, Appl
	125	49.4	4.9	12237	12	US-10-311-455-2332 Sequence 2332, Ap
	126	49.2	4.9	572	13	US-10-027-632-321875 Sequence 321875,
c	127	49.2	4.9	894	14	US-10-156-761-3126 Sequence 3126, Ap
	128	49.2	4.9	249487	13	US-10-026-188-3 Sequence 3, Appli
	129	49.2	4.9	9025608	14	US-10-156-761-1 Sequence 1, Appli
c	130	49	4.9	1446	14	US-10-156-761-4465 Sequence 4465, Ap
c	131	48.8	4.9	210	10	US-09-933-797-404 Sequence 404, App
	132	48.8	4.9	239	13	US-10-027-632-283353 Sequence 283353,
c	133	48.8	4.9	1682	9	US-09-755-325-7 Sequence 7, Appli
c	134	48.8	4.9	2561	10	US-09-976-740-48 Sequence 48, Appl
c	135	48.8	4.9	2561	13	US-10-023-529-48 Sequence 48, Appl
c	136	48.8	4.9	2561	13	US-10-023-523-48 Sequence 48, Appl
	137	48.6	4.9	440	14	US-10-184-644-202 Sequence 202, App
	138	48.6	4.9	440	14	US-10-184-634-202 Sequence 202, App
c	139	48.6	4.9	1973	13	US-10-027-632-97352 Sequence 97352, A
	140	48.6	4.9	3503	9	US-09-823-038A-58 Sequence 58, Appl
c	141	48.6	4.9	263744	12	US-10-229-834A-6 Sequence 6, Appli
c	142	48.4	4.8	3038	13	US-10-027-632-115184 Sequence 115184,
	143	48.2	4.8	5452	12	US-10-017-161-1481 Sequence 1481, Ap
c	144	48.2	4.8	397658	10	US-09-813-320-3 Sequence 3, Appli
	145	48	4.8	496	11	US-09-918-995-11733 Sequence 11733, A
	146	48	4.8	157875	11	US-09-935-464-1 Sequence 1, Appli
	147	48	4.8	157875	14	US-10-125-835-1 Sequence 1, Appli
	148	47.8	4.8	486	13	US-10-027-632-135836 Sequence 135836,
	149	47.8	4.8	2487	13	US-10-027-632-103365 Sequence 103365,
c	150	47.8	4.8	1223197	13	US-10-027-632-179264 Sequence 179264,
	151	47.6	4.8	14776	10	US-09-880-107-2088 Sequence 2088, Ap
	152	47.6	4.8	85548	14	US-10-175-523-75 Sequence 75, Appl
c	153	47.6	4.8	225883	14	US-10-175-523-57 Sequence 57, Appl
	154	47.6	4.8	659158	10	US-09-771-208-20 Sequence 20, Appl
	155	47.4	4.7	605	13	US-10-027-632-271134 Sequence 271134,
	156	47.4	4.7	702	13	US-10-027-632-126555 Sequence 126555,
	157	47.4	4.7	1218	13	US-10-027-632-256923 Sequence 256923,
	158	47.4	4.7	2436	13	US-10-027-632-256922 Sequence 256922,
c	159	47.4	4.7	4139	10	US-09-964-824A-105 Sequence 105, App
c	160	47.4	4.7	4139	10	US-09-964-824A-578 Sequence 578, App
c	161	47.4	4.7	4139	10	US-09-864-864-334 Sequence 334, App
c	162	47.4	4.7	4139	10	US-09-880-107-2121 Sequence 2121, Ap
c	163	47.4	4.7	4139	14	US-10-171-311-157 Sequence 157, App
c	164	47.4	4.7	4139	14	US-10-177-293-310 Sequence 310, App
	165	47.4	4.7	6419	12	US-10-311-455-239 Sequence 239, App
c	166	47.2	4.7	576	13	US-10-027-632-272310 Sequence 272310,
c	167	47.2	4.7	1990	14	US-10-198-846-9710 Sequence 9710, Ap
c	168	47.2	4.7	2678	12	US-10-017-161-1647 Sequence 1647, Ap
	169	47	4.7	628	13	US-10-027-632-11159 Sequence 11159, A
	170	47	4.7	628	13	US-10-027-632-11160 Sequence 11160, A
	171	47	4.7	628	13	US-10-027-632-143978 Sequence 143978,
	172	47	4.7	130427	14	US-10-175-523-87 Sequence 87, Appl
c	173	46.8	4.7	585	13	US-10-027-632-319580 Sequence 319580,
c	174	46.8	4.7	585	13	US-10-027-632-319581 Sequence 319581,
c	175	46.8	4.7	585	13	US-10-027-632-319582 Sequence 319582,
	176	46.8	4.7	666	13	US-10-027-632-28781 Sequence 28781, A
c	177	46.8	4.7	672	13	US-10-027-632-106237 Sequence 106237,
	178	46.8	4.7	702	13	US-10-027-632-110828 Sequence 110828,
	179	46.8	4.7	702	13	US-10-027-632-110829 Sequence 110829,
	180	46.8	4.7	702	13	US-10-027-632-110830 Sequence 110830,
c	181	46.8	4.7	819	13	US-10-027-632-130210 Sequence 130210,
	182	46.8	4.7	922	12	US-10-017-161-1435 Sequence 1435, Ap
	183	46.8	4.7	38186	11	US-09-373-658-38 Sequence 38, Appl
c	184	46.8	4.7	118067	15	US-10-081-327-32 Sequence 32, Appl
c	185	46.6	4.7	577	13	US-10-027-632-231482 Sequence 231482,
	186	46.6	4.7	1393	9	US-09-454-540-1 Sequence 1, Appli
	187	46.6	4.7	1393	10	US-09-859-894A-1 Sequence 1, Appli
	188	46.6	4.7	1393	10	US-09-841-730-24 Sequence 24, Appl
c	189	46.6	4.7	2430	14	US-10-156-761-2531 Sequence 2531, Ap
	190	46.6	4.7	43058	10	US-09-954-456-292 Sequence 292, App
	191	46.6	4.7	43058	10	US-09-954-456-529 Sequence 529, App
	192	46.6	4.7	43058	10	US-09-880-107-3950 Sequence 3950, Ap
c	193	46.6	4.7	152331	13	US-10-095-407-16 Sequence 16, Appl
c	194	46.6	4.7	9025608	14	US-10-156-761-1 Sequence 1, Appli
	195	46.4	4.6	573	9	US-09-864-761-9961 Sequence 9961, Ap
	196	46.4	4.6	573	9	US-09-864-761-26463 Sequence 26463, A
c	197	46.4	4.6	672	13	US-10-027-632-137449 Sequence 137449,
c	198	46.4	4.6	672	13	US-10-027-632-137450 Sequence 137450,
c	199	46.4	4.6	702	13	US-10-027-632-201205 Sequence 201205,
c	200	46.4	4.6	1614	10	US-09-976-740-45 Sequence 45, Appl
c	201	46.4	4.6	1614	13	US-10-023-529-45 Sequence 45, Appl
c	202	46.4	4.6	1614	13	US-10-023-523-45 Sequence 45, Appl
c	203	46.4	4.6	11729	12	US-10-311-455-867 Sequence 867, App
c	204	46.4	4.6	12425	10	US-09-976-740-50 Sequence 50, Appl
c	205	46.4	4.6	12425	13	US-10-023-529-50 Sequence 50, Appl
c	206	46.4	4.6	12425	13	US-10-023-523-50 Sequence 50, Appl
c	207	46.4	4.6	2140405	13	US-10-027-632-76212 Sequence 76212, A
c	208	46.2	4.6	564	13	US-10-027-632-274098 Sequence 274098,
c	209	46.2	4.6	693	10	US-09-764-877-2421 Sequence 2421, Ap
	210	46.2	4.6	80557	15	US-10-080-170-647 Sequence 647, App
c	211	46	4.6	607	13	US-10-027-632-42357 Sequence 42357, A
c	212	46	4.6	607	13	US-10-027-632-42358 Sequence 42358, A
c	213	46	4.6	607	13	US-10-027-632-63776 Sequence 63776, A
c	214	46	4.6	607	13	US-10-027-632-63777 Sequence 63777, A
	215	46	4.6	644	14	US-10-198-846-7696 Sequence 7696, Ap
c	216	46	4.6	2307	11	US-09-893-519A-87 Sequence 87, Appl
c	217	46	4.6	42999	10	US-09-799-462A-17 Sequence 17, Appl
c	218	46	4.6	42999	11	US-09-836-911A-17 Sequence 17, Appl
c	219	46	4.6	42999	13	US-10-125-767-17 Sequence 17, Appl
c	220	46	4.6	42999	14	US-10-151-081-17 Sequence 17, Appl
c	221	46	4.6	42999	14	US-10-287-313-17 Sequence 17, Appl
c	222	46	4.6	42999	14	US-10-219-694-17 Sequence 17, Appl
	223	46	4.6	110096	10	US-09-880-107-1542 Sequence 1542, Ap
	224	45.8	4.6	745	13	US-10-027-632-144367 Sequence 144367,
	225	45.8	4.6	931	14	US-10-198-846-8585 Sequence 8585, Ap
c	226	45.8	4.6	32082	11	US-09-764-891-9679 Sequence 9679, Ap
c	227	45.8	4.6	41936	10	US-09-967-768A-116 Sequence 116, App
	228	45.8	4.6	90541	9	US-09-759-359A-3 Sequence 3, Appli
c	229	45.8	4.6	145831	10	US-09-969-708-79 Sequence 79, Appl
c	230	45.8	4.6	145831	10	US-09-954-456-2116 Sequence 2116, Ap
c	231	45.6	4.6	1496	10	US-09-969-708-132 Sequence 132, App
c	232	45.6	4.6	1496	14	US-10-171-581-120 Sequence 120, App
c	233	45.6	4.6	1496	14	US-10-006-430-3 Sequence 3, Appli
c	234	45.4	4.5	563	13	US-10-027-632-210174 Sequence 210174,
c	235	45.4	4.5	563	13	US-10-027-632-210175 Sequence 210175,

	236	49.2	4.5	1758	12	US-10-101-510-450 Sequence 450, App
	237	45.2	4.5	2826	14	US-10-225-567A-43 Sequence 43, Appl
	238	45.2	4.5	3066	14	US-10-121-988-152 Sequence 152, App
	239	45	4.5	499	13	US-10-027-632-195772 Sequence 195772,
c	240	45	4.5	564	13	US-10-027-632-274099 Sequence 274099,
c	241	45	4.5	861	13	US-10-027-632-154904 Sequence 154904,
c	242	45	4.5	1647	13	US-10-027-632-154905 Sequence 154905,
	243	45	4.5	1747	12	US-10-017-161-2371 Sequence 2371, Ap
	244	45	4.5	1747	12	US-10-017-161-2379 Sequence 2379, Ap
	245	45	4.5	6453	12	US-09-960-706-618 Sequence 618, App
	246	45	4.5	35425	12	US-10-017-161-2429 Sequence 2429, Ap
	247	45	4.5	173808	13	US-10-003-806-10 Sequence 10, Appl
	248	45	4.5	1601042	13	US-10-027-632-59064 Sequence 59064, A
	249	44.8	4.5	468	11	US-09-918-995-11647 Sequence 11647, A
	250	44.8	4.5	480	13	US-10-027-632-72218 Sequence 72218, A
	251	44.8	4.5	480	13	US-10-027-632-72219 Sequence 72219, A
	252	44.8	4.5	480	13	US-10-027-632-72220 Sequence 72220, A
	253	44.8	4.5	480	13	US-10-027-632-72801 Sequence 72801, A
	254	44.8	4.5	480	13	US-10-027-632-72802 Sequence 72802, A
	255	44.8	4.5	480	13	US-10-027-632-108633 Sequence 108633,
	256	44.8	4.5	480	13	US-10-027-632-108634 Sequence 108634,
	257	44.8	4.5	480	13	US-10-027-632-108635 Sequence 108635,
c	258	44.8	4.5	1166	13	US-10-027-632-251519 Sequence 251519,
c	259	44.8	4.5	1166	13	US-10-027-632-251520 Sequence 251520,
c	260	44.8	4.5	2892	14	US-10-175-523-103 Sequence 103, App
c	261	44.6	4.5	19994	12	US-10-017-161-729 Sequence 729, App
c	262	44.4	4.4	362	13	US-10-027-632-285908 Sequence 285908,
	263	44.4	4.4	594	12	US-10-140-472-10 Sequence 10, Appl
	264	44.4	4.4	594	12	US-10-141-761-10 Sequence 10, Appl
	265	44.4	4.4	594	14	US-10-123-155-10 Sequence 10, Appl
	266	44.4	4.4	594	15	US-10-146-731-10 Sequence 10, Appl
	267	44.4	4.4	692	13	US-10-027-632-96253 Sequence 96253, A
	268	44.4	4.4	692	13	US-10-027-632-306160 Sequence 306160,
	269	44.4	4.4	931	13	US-10-027-632-120060 Sequence 120060,
	270	44.4	4.4	931	13	US-10-027-632-120061 Sequence 120061,
	271	44.4	4.4	931	13	US-10-027-632-120062 Sequence 120062,
c	272	44.4	4.4	1647	14	US-10-156-761-5744 Sequence 5744, Ap
c	273	44.4	4.4	1854	10	US-09-912-599-2 Sequence 2, Appli
c	274	44.4	4.4	186510	12	US-10-043-715-1 Sequence 1, Appli
	275	44.2	4.4	227	12	US-10-084-843-165 Sequence 165, App
	276	44.2	4.4	227	12	US-10-193-002-160 Sequence 160, App
	277	44.2	4.4	400	12	US-10-084-843-179 Sequence 179, App
	278	44.2	4.4	400	12	US-10-193-002-174 Sequence 174, App
	279	44.2	4.4	2137	14	US-10-198-846-11433 Sequence 11433, A
	280	44.2	4.4	176373	13	US-10-095-407-17 Sequence 17, Appl
c	281	44	4.4	506	13	US-10-027-632-196339 Sequence 196339,
	282	44	4.4	652	13	US-10-027-632-247562 Sequence 247562,
	283	44	4.4	652	13	US-10-027-632-247563 Sequence 247563,
	284	44	4.4	652	13	US-10-027-632-247564 Sequence 247564,
c	285	44	4.4	702	13	US-10-027-632-110828 Sequence 110828,
c	286	44	4.4	702	13	US-10-027-632-110829 Sequence 110829,
c	287	44	4.4	702	13	US-10-027-632-110830 Sequence 110830,
c	288	44	4.4	771	13	US-10-027-632-133833 Sequence 133833,
c	289	44	4.4	771	13	US-10-027-632-133834 Sequence 133834,
	290	44	4.4	777	13	US-10-027-632-160333 Sequence 160333,
c	291	44	4.4	1424	9	US-09-729-674-71 Sequence 71, Appl
c	292	44	4.4	2157	13	US-10-027-632-97717 Sequence 97717, A
c	293	44	4.4	4739	10	US-09-954-531-586 Sequence 586, App
c	294	44	4.4	33719	15	US-10-080-170-651 Sequence 651, App
c	295	44	4.4	35465	14	US-10-161-572-6 Sequence 6, Appli
c	296	44	4.4	36991	14	US-10-161-572-8 Sequence 8, Appli
c	297	44	4.4	123192	14	US-10-175-523-71 Sequence 71, Appl
	298	43.8	4.4	587	13	US-10-027-632-196101 Sequence 196101,
	299	43.8	4.4	587	13	US-10-027-632-196102 Sequence 196102,
	300	43.8	4.4	587	13	US-10-027-632-196103 Sequence 196103,

ALIGNMENTS

RESULT 1
US-09-864-761-26582
; Sequence 26582, Application US/09864761

; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FC
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 26582
; LENGTH: 576
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000111.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
; OTHER INFORMATION: SWISSPROT HIT: P44836, EVALUE 1.00e-12
US-09-864-761-26582

Query Match 19.3%; Score 193.4; DB 9; Length 576;
Best Local Similarity 59.0%; Pred. No. 1.2e-34;
Matches 332; Conservative 0; Mismatches 231; Indels 0; Gaps 0;

```
Qy        303 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACA 362
              || || ||| |  |       |   || || ||| | ||| |  |||||| |  |  |
Db          1 GCTTGGTTGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGCTTGGTTGGCTGGTT 60

Qy        363 GTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 422
              |   || || | ||||| | |   || || | | | |  ||||||||||||| ||| || |
Db         61 GCTTGGTTGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTGCTTGCTT 120

Qy        423 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 482
```

```
Db        121 GCTTGCTTGCTTGCTTGCTTGGCTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGCTT 180

Qy        483 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 542

Db        181 GCTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTT 240

Qy        543 GCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATG 602

Db        241 GCTTGCTTCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGGTTGGTTGGTTGGTT 300

Qy        603 GCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 662

Db        301 GGTTGCTTGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGGTTGTTTGGTTGCTT 360

Qy        663 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTG 722

Db        361 GGTTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGCTTGCTTTGTTGCTT 420

Qy        723 GGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 782

Db        421 GCTTGGTTCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTT 480

Qy        783 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTT 842

Db        481 CGTTGATTGCTTGGTTGGTTGGTTGCTTGGTTGGTTGGCTGGTTCATTGGTTGGTTGGTT 540

Qy        843 GCTAGTGGGGCCGCTTGCCTGCT 865

Db        541 GCTTGGTTGTTTGGTTGGTTGGT 563
```

```
RESULT 2
US-09-864-761-9605
; Sequence 9605, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 9605
;   LENGTH: 555
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AP000111.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
US-09-864-761-9605

  Query Match             17.4%;  Score 174.6;  DB 9;  Length 555;
  Best Local Similarity   58.9%;  Pred. No. 1.9e-30;
  Matches  300;  Conservative    0;  Mismatches  209;  Indels    0;  Gaps    0;

Qy        337 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGC 396

Db         46 TTGGTTGACTGGTTGCTTGGTTGGTTGGTTCGTTGGTTGCTTGGTTGGTTGCTTGGTTGC 105

Qy        397 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 456

Db        106 TTGGTTGGCTGGTTGCTTGGTTGTTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGC 165

Qy        457 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 516

Db        166 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGGCTGGTTGGTTGGTTGGTTGCTTGG 225

Qy        517 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGC 576

Db        226 TTGGTTGGTTGCTTGCTTGCTTGGTTGCTTGCTTCATTGGTTGCTTGGTTGCTTGCTTCA 285

Qy        577 TAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGG 636

Db        286 TTGGTTGCTTGGTTGCTTGCTTCGTTGCTTGCTTGGTTGGTTGCTTGGTTGGTTGCTTGG 345

Qy        637 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 696

Db        346 TTGGTTGGTTGGTTGGTTGCTTGATTGGTTGCTTGGTTGCTTGGTTGGTTGGTTGGTTGG 405

Qy        697 CTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGG 756

Db        406 TTGTTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGGTTGGTTGGTTGGTTGCTTGC 465

Qy        757 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 816

Db        466 TTGCTTTGTTGCTTGCTTGGTTCGTTGGTTCGTTGGTTGGTTGGTTGGTTGGTTGGTTGG 525

Qy        817 CTGCCTGCTGGCTTGCTTGCTTGCTTGCT 845

Db        526 TTGGTTGGTTGGTTCGTTGATTGCTTGGT 554
```

```
RESULT 3
US-10-017-161-1345
; Sequence 1345, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
```

```
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1345
;   LENGTH: 1360
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1360)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1160)
US-10-017-161-1345

  Query Match             11.1%;  Score 111.4;  DB 12;  Length 1360;
  Best Local Similarity   49.9%;  Pred. No. 2.2e-16;
  Matches  310;  Conservative    0;  Mismatches  306;  Indels    5;  Gaps    1;

Qy        219 GGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTG 278
              | ||||||||||||  ||  |||||| |||     | | ||  ||| |  |||  ||
Db        328 GTCTGCCTGCCTGCCTGCCTGCCTGCCTGTCTGTCTGTCTGCCCGTCTGCCTGCCC 387

Qy        279 GCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 338
                ||||  ||       | | ||  | |||  ||   |   | |  |  ||||
Db        388 ATCTGCCTGTCTGTCTGCCTGTCCGTCTGTCTGTCCATCTGTCCATCTGCCTATCCATCT 447

Qy        339 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 398
              ||  |||||  |||       | | |   | |||   || ||     |  | |      ||
Db        448 GCCTGCCTGTCTGCCTGTCCGTCTGCCTGTCTGTCTGCCTGTCCATCTGTCCATCTGCCT 507

Qy        399 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 458
                |   |||  |||  |   | | ||  |||  |||  ||  | |||  ||||  ||  |||
Db        508 ATCCATCTGCCTGCCTGTCTGTCGGCCTGCCTGCCTGCCTGTCTGTCTGCTGCCTGTCTG 567

Qy        459 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 518
               ||  ||  ||  ||     |||||   |  |||  ||  | | |  |||  ||  ||
Db        568 TCCGTCTGCCTGTCT-----GCCTGTCCGTCTGCCTGCCTGTCCGTCTGCCTGTCCGTCT 622

Qy        519 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTA 578
              ||||||  |||  ||   |   ||||||  ||        | | ||  | ||||  ||   |
Db        623 GCCTGCCTGCCTGTCTGTCTGCCTGCCTGTCTGCCTGCCTGTCCGTCTGCCTGTCCGTCT 682

Qy        579 AGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 638
                | | ||  |||||||||   ||  ||   |  | ||||  | |||   |   | |
Db        683 GCCTGCCTGTCTGCCTGCCTGTCTGCCTGTCTGCCCGTCTGCCTGTCTGTCTGCCTGTCC 742

Qy        639 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 698
              |   | |||   ||  | |||     ||||   |    ||||  | ||    ||  | |
Db        743 GTCTGCCTGTCTGTCCGTCTGTCCATCTGCCTATCCATCTGCCTGCCTATCTGTCTGTCC 802

Qy        699 CCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCT 758
                | ||||  ||  | ||||  | ||||  ||  | ||  | |||    |  | ||
Db        803 GTCTGCCTGCCTGTCTGTCTGCCTGTCTGCCTGTCTGTCTGCCTGTCTGTCCATCTGCCT 862

Qy        759 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 818
                 || |  |   | |||   | |||   ||  | |||   | ||||  |   | ||
Db        863 ATCCATCTACCTGCCTGCCTGTCTGCCTGTCTGTCTGCCTGTCTGTCTGCCTGCCTGTCT 922

Qy        819 GCCTGCTGGCTTGCTTGCTTG 839
              | |||   |  || |||||||
Db        923 GTCTGTCTGTCTGGTTGCTTG 943


RESULT 4
US-10-017-161-2117
; Sequence 2117, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2117
;   LENGTH: 1672
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1672)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1472)
US-10-017-161-2117

  Query Match             9.5%;  Score 94.6;  DB 12;  Length 1672;
  Best Local Similarity   47.0%;  Pred. No. 1.2e-12;
  Matches  325;  Conservative    0;  Mismatches  364;  Indels    2;  Gaps    1;

Qy        189 GCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGC 248
              |||||||| ||| ||| |     |  | |     ||  ||| |||  |   |  |||||||
Db        536 GCGTGTGCTTGCAGGTGTGCGTGCTTGCAGCTGTGTGTGCTTGCAGGTGTGTGTGCTTGC 595

Qy        249 TG--AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 306
              ||    | |   | ||   ||  |||| || ||   |  ||||    ||   | | || |
Db        596 TGCTGTGCGCTTGCAGCTGTGTGTGCTTGCAGGTGCGTGTGCTTGCAGGTGTGCATGCTT 655

Qy        307 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAG 366
              ||         | |  | |||  ||  |||| ||||   ||        ||  |   ||
Db        656 GCAGCTGTGTGTGCTTGCAGGTGTGTGTGCTTGCTGCTGTGTGCTTGCAGCTGTGTGTGC 715

Qy        367 GGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 426
                        |  ||     |  | ||   ||| ||     ||  |||| || ||  ||| |
Db        716 TTGCAGGTGTGTGTGCTTGCTGGTGTGTGCTTGCAGCTGTGTGTGCTTGCAGGTGTGCGT 775

Qy        427 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 486
              ||||||      |  |||||||  |  ||   || |||      ||   ||| ||  |  |
Db        776 GCTTGCAGCTGTGTGTGCTTGCAGGTGCGTGTGCTTGCAGCTGTGTGTGCTTGCAGGTGT 835

Qy        487 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 546
              |  |||| || ||   |  |||||||      |  |||||||      |   || |||
Db        836 GTGTGCTTGCAGGTGCGTGTGCTTGCAGCTGTGTGTGCTTGCAGCTGTGTGTGCTTGCAG 895

Qy        547 GCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTT 606
                  | |   | ||   |  |||||||    | |   |   || |  ||   ||||
Db        896 CTGTGTGTGCTTGCTGGTGCGTGTGCTTGCAGCTGTGTGTGCTTGCAGGTGTGTGTGCTT 955

Qy        607 GATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 666
              |    | |  ||||  |   |   ||  |  ||| |   |||  |   |    | |
Db        956 GCAGCTGTGTGTGCTTGCAGGTGCGTGTGCTTGCAGCTGTGCGTGCTTGCAGCTGTGCGT 1015

Qy        667 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCT 726
              ||| | ||  ||  |  |||  ||  |  | || |   |  ||  |  ||| |  |
Db       1016 GCTTGCAGCTGTGTGTGCTTGCAGGTGTGTGTGCTTGCAGGTGCGTGTGCTTGCAGCTGT 1075

Qy        727 GCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 786
              || |   |   |  | |||| | ||  ||   |  |||   | |  |||| |
Db       1076 GCGTGCTTGCAGCTGTGCGTGCTTGCAGCTGTGTGTGCTTGCAGCTGTGTGTGCTTGCTG 1135

Qy        787 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTA 846
              |  ||  |  ||| ||  ||   |   |   ||  |||||||  |   ||||
Db       1136 GTGCGTGTGCTTGCAGGTGTGTGTGCTAGCAGGTGTGTGTGCTTGCTGGTGCGTGTGCTT 1195

Qy        847 GTGGGGCCGCTTGCCTGCTACTAGGGCTGCT 877
              |  ||  ||  ||| |||   |  |  ||||
```

```
Db       1196 GCTGGTGCGTGTGCTTGCAGGTGCGTGTGCT 1226
```

RESULT 5
US-10-027-632-164203/c
; Sequence 164203, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 164203
; LENGTH: 798
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-164203

Query Match 8.3%; Score 82.6; DB 13; Length 798;
Best Local Similarity 45.6%; Pred. No. 6.5e-10;
Matches 283; Conservative 2; Mismatches 336; Indels 0; Gaps 0;

```
Qy        217 ACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGC 276
Db        683 ATGCATGAGAGCCTGGATGAAAGACTGACTGGAAGGCTGTTTGACAAGCTGGCTGCCTAA 624

Qy        277 TGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGC 336
Db        623 AAGCCTGGCTCAAAAGCTGCTGGTTGAGAGGCTGGCTTAGAGGCTTGCTGAGAAGCTGGC 564

Qy        337 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGC 396
Db        563 TGTATTAAAGGCTTGCCTAGAGCATGGTTGTCAGGCTGGCTGGCTGTCAGGCTGAGCAAG 504

Qy        397 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 456
Db        503 GGTTTATCTGGAAGGCTGGTCAAGATGCAGTCTGGGTGGCTGCCTGGCTGAGAGGCTGGT 444

Qy        457 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 516
Db        443 TGCCTGGGAAGCTGGCTGAGAGGCTGCCTGCCTGCAAGTCTGGCTTAGAGGGTTACTGAG 384

Qy        517 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGC 576
Db        383 AGGCTGGCTGGATTAGAGGCTTGCCTGGAGGCCAGCTAGTAGGTGGGCTGGCTGAGAGGC 324

Qy        577 TAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGG 636
Db        323 TAGCTGCTTGAAAGGGCAGTTGAGAGGCTGGCTGAGAGGCTGGCTGGGAGGCTGTTTGAG 264

Qy        637 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 696
Db        263 AGGCTGGCTKCCTGAGAGACTGGCTCAGAGGCTGCCTGGTTGAGAGGCTGGCTTAGAGRC 204

Qy        697 CTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGG 756
```

```
Db        203 TTCCTGAGAGCCTGGCTGGTTGAGAAGCTTGCTAAGAGACTGGCTGGCAGGTTGGCTGGC 144

Qy        757 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 816
Db        143 TGAGAGATCTGCCAAGGGGGTGGCTTGGAATCTTGCCCAGAGGTGGGCTGGCTCCCTGGT 84

Qy        817 CTGCCTGCTGGCTTGCTTGCT 837
Db         83 TTAGAGGTTACCTCGAAGGCT 63
```

RESULT 6
US-10-027-632-164204/c
; Sequence 164204, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 164204
; LENGTH: 798
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-164204

Query Match 8.3%; Score 82.6; DB 13; Length 798;
Best Local Similarity 45.6%; Pred. No. 6.5e-10;
Matches 283; Conservative 2; Mismatches 336; Indels 0; Gaps 0;

```
Qy        217 ACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGC 276
Db        683 ATGCATGAGAGCCTGGATGAAAGACTGACTGGAAGGCTGTTTGACAAGCTGGCTGCCTAA 624

Qy        277 TGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGC 336
Db        623 AAGCCTGGCTCAAAAGCTGCTGGTTGAGAGGCTGGCTTAGAGGCTTGCTGAGAAGCTGGC 564

Qy        337 TCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGC 396
Db        563 TGTATTAAAGGCTTGCCTAGAGCATGGTTGTCAGGCTGGCTGGCTGTCAGGCTGAGCAAG 504

Qy        397 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 456
Db        503 GGTTTATCTGGAAGGCTGGTCAAGATGCAGTCTGGGTGGCTGCCTGGCTGAGAGGCTGGT 444

Qy        457 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 516
Db        443 TGCCTGGGAAGCTGGCTGAGAGGCTGCCTGCCTGCAAGTCTGGCTTAGAGGGTTACTGAG 384

Qy        517 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGC 576
Db        383 AGGCTGGCTGGATTAGAGGCTTGCCTGGAGGCCAGCTAGTAGGTGGGCTGGCTGAGAGGC 324

Qy        577 TAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGG 636
```

```
Db        323 TAGCTGCTTGAAAGGGCAGTTGAGAGGCTGGCTGAGAGGCTGGCTGGGAGGCTGTTTGAG 264

Qy        637 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 696

Db        263 AGGCTGGCTKCCTGAGAGACTGGCTCAGAGGCTGCCTGGTTGAGAGGCTGGCTTAGAGRC 204

Qy        697 CTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGG 756

Db        203 TTCCTGAGAGCCTGGCTGGTTGAGAAGCTTGCTAAGAGACTGGCTGGCAGGTTGGCTGGC 144

Qy        757 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 816

Db        143 TGAGAGATCTGCCAAGGGGGTGGCTTGGAATCTTGCCCAGAGGTGGGCTGGCTCCCTGGT 84

Qy        817 CTGCCTGCTGGCTTGCTTGCT 837

Db         83 TTAGAGGTTACCTCGAAGGCT 63


RESULT 7
US-10-017-161-1915
; Sequence 1915, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1915
;   LENGTH: 14750
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(14750)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(343)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (5163)..(5441)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (13911)..(14088)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (14443)..(14550)
US-10-017-161-1915

  Query Match             8.1%;  Score 81.4;  DB 12;  Length 14750;
  Best Local Similarity   47.6%;  Pred. No. 8.3e-10;
  Matches  308;  Conservative    0;  Mismatches  331;  Indels    8;  Gaps    2;

Qy        226 TGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCT 285

Db        929 TGCTGGGGTGCTGGGGTGCTAGGATGCTGAGGTTTTGTGATACTGGGTCGCTGAGATGCT 988

Qy        286 TGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 345

Db        989 GGGATACTGGGGTGCTGAAGTGCTGGGGTACTGGGGTGTTGCCATGCTGAAATGCTGGGG 1048

Qy        346 TGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTA-------GGGCTGCTCGCTG 398

Db       1049 TGTTGGGATGCTGAGGTGCTGGGTTGCTGGGATGCTGGGATAGTCCTTGGATGCTGCGGT 1108

Qy        399 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 458

Db       1109 GCTGAGATGCTGGTCTGCTGGTCTGCTGGGGTGTTGGGATGCTGGGATTTTGGGATGCTG 1168

Qy        459 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 518

Db       1169 GAGTGCTGCAGTGCTGCAGTGCTGAGATGCTGGAGTGCTGGGGTGCTGGAATACTGTAGT 1228

Qy        519 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTA 578

Db       1229 ACTGGTGTGCTGGAGCATTGAGATGCTAGGGCACTGGGATGCTAAGGTGCTGAGATGCTG 1288

Qy        579 AGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 638

Db       1289 CAGTGCTGGGGTGTTGGGATGCCGAGGTGTTGGGATGCTTAAGTGCTGGGGTGTTGGGAT 1348

Qy        639 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 698

Db       1349 GCTGGGGTGCTGGAACACTATGGTGCTGGGGTGCTGGAGTGTTGAGATACTGTAGTGGTG 1408

Qy        699 CCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCT 758

Db       1409 GGATGCTTAAGTGCTGGGGTGCTGGG-TGTTGGGATGCCTAGGTGCTGGGGTGCCGAGAT 1467

Qy        759 GCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCT 818

Db       1468 GCTGGAGTACTAGTGTGCTGGGATGCTGAAGTGCTGGGGTCCTGAGATGCTGCAGTGCTA 1527

Qy        819 GCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCT 865

Db       1528 GGGTGCTGAGATTCTGGGCTGCTGGAGTGTTGGGGTGCTGGGATGCT 1574


RESULT 8
US-10-329-960-1
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
; GENERAL INFORMATION:
;  APPLICANT: Fleischmann et al.
;  TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, F
;  TITLE OF INVENTION:  Thereof, and Uses Thereof
;  FILE REFERENCE: PB186P1
;  CURRENT APPLICATION NUMBER: US/10/329,960
;  CURRENT FILING DATE:  2003-01-02
;  PRIOR APPLICATION NUMBER: US 09/643,990
;  PRIOR FILING DATE: 2000-08-23
;  PRIOR APPLICATION NUMBER: US 08/487,429
;  PRIOR FILING DATE: 1995-06-07
;  PRIOR APPLICATION NUMBER: US 08/426,787
;  PRIOR FILING DATE: 1995-04-21
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1830121
;   TYPE: DNA
;   ORGANISM: Haemophilus influenzae
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4747)..(4747)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9921)..(9921)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (10150)..(10150)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (29298)..(29298)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
```

; NAME/KEY: misc_feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145942)..(145942)

```
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (147197)..(147197)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (150841)..(150841)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (152500)..(152500)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (152530)..(152530)

  Query Match             7.7%;  Score 77.4;  DB 14;  Length 1830121;
  Best Local Similarity   84.5%;  Pred. No. 3.5e-09;
  Matches   87;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy        449 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 508
              ||||| ||| || |||||||||||| ||| || |||||||||||| ||| || |||||||
Db    1543129 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGC 1543188

Qy        509 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGA 551
              ||||| ||| || |||||||||||| ||| || ||||||| ||
Db    1543189 TTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCAGA 1543231


RESULT 9
US-09-799-462A-17
; Sequence 17, Application US/09799462A
; Patent No. US20020160970A1
;   GENERAL INFORMATION:
;        APPLICANT: Hadlaczky, Gyula
;                   Szalay, Aladar
;        TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
;                            AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
;        NUMBER OF SEQUENCES: 34
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Heller Ehrman White & McAuliffe
;             STREET: 4250 Executive Square, 7th Floor
;             CITY: La Jolla
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 92037
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ Version 1.5
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/799,462A
;             FILING DATE: 10-Sep-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/835,682
;             FILING DATE: 10-APR-1997
;             APPLICATION NUMBER: 08/695,191
;             FILING DATE: 07-AUG-1996
;             APPLICATION NUMBER: 08/682,080
;             FILING DATE: 15-JUL-1996
;             APPLICATION NUMBER: 08/629,822
;             FILING DATE: 10-APR-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Seidman, Stephanie L
;             REGISTRATION NUMBER: 33,779
;             REFERENCE/DOCKET NUMBER: 24601-402G
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 858-450-8403
;             TELEFAX: 858-587-5360
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 17:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 42999 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: Genomic DNA
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        FRAGMENT TYPE: <Unknown>
;        ORIGINAL SOURCE:
;        SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-799-462A-17

  Query Match             7.7%;  Score 77;  DB 10;  Length 42999;
  Best Local Similarity   58.7%;  Pred. No. 6.9e-09;
  Matches  172;  Conservative    0;  Mismatches  115;  Indels    6;  Gaps    2;

Qy        263 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACA 322
              || | |||||| ||| || | | ||||     |    | |   |  ||  |
Db      40286 GCTTTCTTGCTTGCTTGCTTTCGTGCTGTCTTGTTTCTCGATTTCTTTCTTTCTTTTGTT 40345

Qy        323 GTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGC-- 380
              ||||||||||  || || || |||||    | |       |  | || | || | |
Db      40346 TCTTTCCTGCTTGCTTTCTTGCTTGATTGCTTTCGTGCTTTCTTGCTTTCTTGTTTTCTT 40405

Qy        381 ---TAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 437
                |  |   | | ||  ||  | ||||| |||| |  || || | ||||||||| |||
Db      40406 TCTTTCTTTTGTTTCTTTCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCT 40465

Qy        438 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC-TTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
               | ||||| |||| |  || || | ||| | ||  |  ||  | ||||||||| || ||
Db      40466 TTCGTGCTTTCTTGTTTTCTTGCTTTCTTTCTTTTGTTTCTTTCTTGCTTGCTTTCTTGC 40525

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              |  | || || ||||||  || || |||||||||  |  |||  | | ||||||
Db      40526 TTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTTCTTTCTTGCT 40578


RESULT 10
US-09-836-911A-17
; Sequence 17, Application US/09836911A
; Publication No. US20030033617A1
;   GENERAL INFORMATION:
;        APPLICANT: Hadlaczky, Gyula
;                   Szalay, Aladar
;        TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
;                            AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
;        NUMBER OF SEQUENCES: 34
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Heller Ehrman White & McAuliffe
;             STREET: 4350 La Jolla Village Drive, 6th Floor
;             CITY: San Diego
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 92122
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ Version 1.5
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/836,911A
;             FILING DATE: 17-Apr-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/835,682
;             FILING DATE: 10-APR-1997
;             APPLICATION NUMBER: 08/695,191
;             FILING DATE: 07-AUG-1996
;             APPLICATION NUMBER: 08/682,080
;             FILING DATE: 15-JUL-1996
;             APPLICATION NUMBER: 08/629,822
```

```
;            FILING DATE: 10-APR-1996
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Seidman, Stephanie L
;            REGISTRATION NUMBER: 33,779
;            REFERENCE/DOCKET NUMBER: 24601-402I
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 858-450-8403
;            TELEFAX: 858-587-5360
;            TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 17:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 42999 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: Genomic DNA
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       FRAGMENT TYPE: <Unknown>
;       ORIGINAL SOURCE:
;       SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-836-911A-17

  Query Match             7.7%;  Score 77;  DB 11;  Length 42999;
  Best Local Similarity   58.7%;  Pred. No. 6.9e-09;
  Matches  172;  Conservative    0;  Mismatches  115;  Indels    6;  Gaps    2;

Qy        263 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACA 322
              || | |||||| ||| || | | |||||    |    | |   |  ||  |
Db      40286 GCTTTCTTGCTTGCTTGCTTTCGTGCTGTCTTGTTTCTCGATTTCTTTCTTTCTTTTGTT 40345

Qy        323 GTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGC-- 380
              ||||||||||  || || || |||||   | |       |  | || | || |
Db      40346 TCTTTCCTGCTTGCTTTCTTGCTTGATTGCTTTCGTGCTTTCTTGCTTTCTTGTTTTCTT 40405

Qy        381 ---TAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 437
                |  |   | | ||  ||  | ||||| |||| |  || || | |||||||||| |||
Db      40406 TCTTTCTTTTGTTTCTTTCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCT 40465

Qy        438 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC-TTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
              | ||||| |||| |  || || | ||| | ||  |  ||  | |||||||||| || ||
Db      40466 TTCGTGCTTTCTTGTTTTCTTGCTTTCTTTCTTTTGTTTCTTTCTTGCTTGCTTTCTTGC 40525

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              |  | || || ||||||  || || |||||||||| |  |||  | | ||||||
Db      40526 TTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTTCTTTCTTGCT 40578


RESULT 11
US-10-125-767-17
; Sequence 17, Application US/10125767
; Publication No. US20020160410A1
;   GENERAL INFORMATION:
;       APPLICANT: Hadlaczky, Gyula
;       TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND
;                           METHODS
;                           FOR PREPARING ARTIFICIAL CHROMOSOMES
;       NUMBER OF SEQUENCES: 34
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Heller Ehrman White & McAuliffe LLP
;            STREET: 4350 La Jolla Village Drive, 7th Floor
;            CITY: San Diego
;            STATE: CA
;            COUNTRY: USA
;            ZIP: 92122
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: DOS
;            SOFTWARE: FastSEQ Version 1.5
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/125,767
;            FILING DATE: 17-Apr-2002
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/724,693
;            FILING DATE: 28-NOV-2000
;            APPLICATION NUMBER: 08/835,682
;            FILING DATE: 10-APR-1997
;            APPLICATION NUMBER: 08/695,191
;            FILING DATE: 07-AUG-1996
;            APPLICATION NUMBER: 08/682,080
;            FILING DATE: 15-JUL-1996
;            APPLICATION NUMBER: 08/629,822
;            FILING DATE: 10-APR-1996
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Seidman, Stephanie L
;            REGISTRATION NUMBER: 33,779
;            REFERENCE/DOCKET NUMBER: 24601-402J
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 858-450-8403
;            TELEFAX: 858-587-5360
;            TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 17:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 42999 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: Genomic DNA
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       FRAGMENT TYPE: <Unknown>
;       ORIGINAL SOURCE:
;       SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-125-767-17

  Query Match             7.7%;  Score 77;  DB 13;  Length 42999;
  Best Local Similarity   58.7%;  Pred. No. 6.9e-09;
  Matches  172;  Conservative    0;  Mismatches  115;  Indels    6;  Gaps    2;

Qy        263 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACA 322
              || | |||||| ||| || | | |||||    |    | |   |  ||  |
Db      40286 GCTTTCTTGCTTGCTTGCTTTCGTGCTGTCTTGTTTCTCGATTTCTTTCTTTCTTTTGTT 40345

Qy        323 GTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGC-- 380
              ||||||||||  || || || |||||   | |       |  | || | || |
Db      40346 TCTTTCCTGCTTGCTTTCTTGCTTGATTGCTTTCGTGCTTTCTTGCTTTCTTGTTTTCTT 40405

Qy        381 ---TAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 437
                |  |   | | ||  ||  | ||||| |||| |  || || | |||||||||| |||
Db      40406 TCTTTCTTTTGTTTCTTTCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCT 40465

Qy        438 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC-TTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
              | ||||| |||| |  || || | ||| | ||  |  ||  | |||||||||| || ||
Db      40466 TTCGTGCTTTCTTGTTTTCTTGCTTTCTTTCTTTTGTTTCTTTCTTGCTTGCTTTCTTGC 40525

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              |  | || || ||||||  || || |||||||||| |  |||  | | ||||||
Db      40526 TTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTTCTTTCTTGCT 40578


RESULT 12
US-10-151-081-17
; Sequence 17, Application US/10151081
; Publication No. US20030083293A1
;   GENERAL INFORMATION:
;       APPLICANT: Hadlaczky, Gyula
;                  Szalay, Aladar
;       TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
;                           AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
;       NUMBER OF SEQUENCES: 34
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Heller Ehrman White & McAuliffe
;            STREET: 4250 Executive Square, 7th Floor
;            CITY: La Jolla
;            STATE: CA
```

```
;          COUNTRY: USA
;          ZIP: 92037
;        COMPUTER READABLE FORM:
;          MEDIUM TYPE: Diskette
;          COMPUTER: IBM Compatible
;          OPERATING SYSTEM: DOS
;          SOFTWARE: FastSEQ Version 1.5
;        CURRENT APPLICATION DATA:
;          APPLICATION NUMBER: US/10/151,081
;          FILING DATE: 16-May-2002
;          CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;          APPLICATION NUMBER: 09/799,462
;          FILING DATE: HEREWITH  05-MAR-2001
;          APPLICATION NUMBER: 09/724,693
;          FILING DATE: HEREWITH  28-NOV-2000
;          APPLICATION NUMBER: 08/835,682
;          FILING DATE: 10-APR-1997
;          APPLICATION NUMBER: 08/695,191
;          FILING DATE: 07-AUG-1996
;          APPLICATION NUMBER: 08/682,080
;          FILING DATE: 15-JUL-1996
;          APPLICATION NUMBER: 08/629,822
;          FILING DATE: 10-APR-1996
;        ATTORNEY/AGENT INFORMATION:
;          NAME: Seidman, Stephanie L
;          REGISTRATION NUMBER: 33,779
;          REFERENCE/DOCKET NUMBER: 24601-402L
;        TELECOMMUNICATION INFORMATION:
;          TELEPHONE: 858-450-8403
;          TELEFAX: 858-587-5360
;          TELEX: <Unknown>
;     INFORMATION FOR SEQ ID NO: 17:
;        SEQUENCE CHARACTERISTICS:
;          LENGTH: 42999 base pairs
;          TYPE: nucleic acid
;          STRANDEDNESS: single
;          TOPOLOGY: linear
;        MOLECULE TYPE: Genomic DNA
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        FRAGMENT TYPE: <Unknown>
;        ORIGINAL SOURCE:
;        SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-151-081-17

  Query Match             7.7%;  Score 77;  DB 14;  Length 42999;
  Best Local Similarity   58.7%;  Pred. No. 6.9e-09;
  Matches  172;  Conservative    0;  Mismatches  115;  Indels    6;  Gaps    2;

Qy        263 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACA 322
              || | |||||| ||| || | | |||||    |    | |    |  ||  |
Db      40286 GCTTTCTTGCTTGCTTGCTTTCGTGCTGTCTTGTTTCTCGATTTCTTTCTTTCTTTTGTT 40345

Qy        323 GTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGC-- 380
              ||||||||||  || || || |||||   | |       |  | || | || |
Db      40346 TCTTTCCTGCTTGCTTTCTTGCTTGATTGCTTTCGTGCTTTCTTGCTTTCTTGTTTTCTT 40405

Qy        381 ---TAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 437
                |  |   | | ||  ||  | ||||| |||| |  || || | ||||||||| |||
Db      40406 TCTTTCTTTTGTTTCTTTCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCT 40465

Qy        438 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC-TTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
                | ||||| |||| |  || || | ||| | ||  |  ||  | ||||||||| || ||
Db      40466 TTCGTGCTTTCTTGTTTTCTTGCTTTCTTTCTTTTGTTTCTTTCTTGCTTGCTTTCTTGC 40525

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              |  | || || ||||||  || || |||||||||  |  |||  | | ||||||
Db      40526 TTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTTCTTTCTTGCT 40578


RESULT 13
US-10-287-313-17
; Sequence 17, Application US/10287313
; Publication No. US20030101480A1
;   GENERAL INFORMATION:
;        APPLICANT: Hadlaczky, Gyula
;                   Szalay, Aladar
;        TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
;                            AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
;        NUMBER OF SEQUENCES: 34
;        CORRESPONDENCE ADDRESS:
;          ADDRESSEE: Heller Ehrman White & McAuliffe
;          STREET: 4250 Executive Square, 7th Floor
;          CITY: La Jolla
;          STATE: CA
;          COUNTRY: USA
;          ZIP: 92037
;        COMPUTER READABLE FORM:
;          MEDIUM TYPE: Diskette
;          COMPUTER: IBM Compatible
;          OPERATING SYSTEM: DOS
;          SOFTWARE: FastSEQ Version 1.5
;        CURRENT APPLICATION DATA:
;          APPLICATION NUMBER: US/10/287,313
;          FILING DATE: 01-No. US20030101480A1-2002
;          CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;          APPLICATION NUMBER: 09/724,726
;          FILING DATE: 28-NOV-2000
;          APPLICATION NUMBER: 08/835,682
;          FILING DATE: 10-APR-1997
;          APPLICATION NUMBER: 08/695,191
;          FILING DATE: 07-AUG-1996
;          APPLICATION NUMBER: 08/682,080
;          FILING DATE: 15-JUL-1996
;          APPLICATION NUMBER: 08/629,822
;          FILING DATE: 10-APR-1996
;        ATTORNEY/AGENT INFORMATION:
;          NAME: Seidman, Stephanie L
;          REGISTRATION NUMBER: 33,779
;          REFERENCE/DOCKET NUMBER: 6869-402N
;        TELECOMMUNICATION INFORMATION:
;          TELEPHONE: 858-450-8403
;          TELEFAX: 858-587-5360
;          TELEX: <Unknown>
;     INFORMATION FOR SEQ ID NO: 17:
;        SEQUENCE CHARACTERISTICS:
;          LENGTH: 42999 base pairs
;          TYPE: nucleic acid
;          STRANDEDNESS: single
;          TOPOLOGY: linear
;        MOLECULE TYPE: Genomic DNA
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        FRAGMENT TYPE: <Unknown>
;        ORIGINAL SOURCE:
;        SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-287-313-17

  Query Match             7.7%;  Score 77;  DB 14;  Length 42999;
  Best Local Similarity   58.7%;  Pred. No. 6.9e-09;
  Matches  172;  Conservative    0;  Mismatches  115;  Indels    6;  Gaps    2;

Qy        263 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACA 322
              || | |||||| ||| || | | |||||    |    | |    |  ||  |
Db      40286 GCTTTCTTGCTTGCTTGCTTTCGTGCTGTCTTGTTTCTCGATTTCTTTCTTTCTTTTGTT 40345

Qy        323 GTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGC-- 380
              ||||||||||  || || || |||||   | |       |  | || | || |
Db      40346 TCTTTCCTGCTTGCTTTCTTGCTTGATTGCTTTCGTGCTTTCTTGCTTTCTTGTTTTCTT 40405

Qy        381 ---TAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 437
                |  |   | | ||  ||  | ||||| |||| |  || || | ||||||||| |||
Db      40406 TCTTTCTTTTGTTTCTTTCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCT 40465
```

```
Qy        438 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC-TTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
              | ||||| |||| |  || || | ||| | ||  |  ||  | ||||||||| || ||
Db      40466 TTCGTGCTTTCTTGTTTTCTTGCTTTCTTTCTTTTGTTTCTTTCTTGCTTGCTTTCTTGC 40525

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              |  | || || ||||||  || || |||||||||  |  |||  | | ||||||
Db      40526 TTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTTCTTTCTTGCT 40578


RESULT 14
US-10-219-694-17
; Sequence 17, Application US/10219694
; Publication No. US20030108914A1
;   GENERAL INFORMATION:
;        APPLICANT: Hadlaczky, Gyula
;        TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS
;                            FOR PREPARING ARTIFICIAL CHROMOSOMES
;        NUMBER OF SEQUENCES: 34
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Heller Ehrman White & McAuliffe LLP
;             STREET: 4350 La Jolla Village Drive, 7th Floor
;             CITY: San Diego
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 92122
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ Version 1.5
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/219,694
;             FILING DATE: 14-Aug-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: <Unknown>
;             FILING DATE: 28-NOV-2000
;             APPLICATION NUMBER: 08/835,682
;             FILING DATE: 10-APR-1997
;             APPLICATION NUMBER: 08/695,191
;             FILING DATE: 07-AUG-1996
;             APPLICATION NUMBER: 08/682,080
;             FILING DATE: 15-JUL-1996
;             APPLICATION NUMBER: 08/629,822
;             FILING DATE: 10-APR-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Seidman, Stephanie L
;             REGISTRATION NUMBER: 33,779
;             REFERENCE/DOCKET NUMBER: 24601-402M
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 858-450-8403
;             TELEFAX: 858-587-5360
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 17:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 42999 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: Genomic DNA
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        FRAGMENT TYPE: <Unknown>
;        ORIGINAL SOURCE:
;        SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-219-694-17

  Query Match             7.7%;  Score 77;  DB 14;  Length 42999;
  Best Local Similarity   58.7%;  Pred. No. 6.9e-09;
  Matches  172;  Conservative    0;  Mismatches  115;  Indels    6;  Gaps    2;

Qy        263 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACA 322
              || | |||||| ||| || | | |||||    |    | |   |  ||  |

Db      40286 GCTTTCTTGCTTGCTTGCTTTCGTGCTGTCTTGTTTCTCGATTTCTTTCTTTCTTTTGTT 40345

Qy        323 GTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGC-- 380
              ||||||||||  || || || |||||   | |       |  | || | || | .
Db      40346 TCTTTCCTGCTTGCTTTCTTGCTTGATTGCTTTCGTGCTTTCTTGCTTTCTTGTTTTCTT 40405

Qy        381 ---TAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 437
                 |  |   | | ||  ||  | ||||| |||| |  || || | ||||||||| |||
Db      40406 TCTTTCTTTTGTTTCTTTCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCT 40465

Qy        438 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC-TTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
              | ||||| |||| |  || || | ||| | ||  |  || | ||||||||| || ||
Db      40466 TTCGTGCTTTCTTGTTTTCTTGCTTTCTTTCTTTTGTTTCTTTCTTGCTTGCTTTCTTGC 40525

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              |  | || || ||||||  || || |||||||||  |  |||  | | ||||||
Db      40526 TTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTTCTTTCTTGCT 40578


RESULT 15
US-09-728-446-369/c
; Sequence 369, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 369
;   LENGTH: 428
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-728-446-369

  Query Match             7.5%;  Score 74.8;  DB 9;  Length 428;
  Best Local Similarity   61.1%;  Pred. No. 3.9e-08;
  Matches  121;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy        377 TTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 436
              ||||| |  ||   | | |||||  |||| | | ||| | ||||  | |||| | | ||
Db        218 TTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGC 159

Qy        437 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
              |||   ||| | ||| | | ||||  | | ||| | | |||||  |||| | | ||| |
Db        158 TGGATGGCTGGGTTGGTGGATGGCTGGATGGCTGGATGGCTGGTTTGCTGGATGGCTGGA 99

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 556
              ||||  | ||| | | | |||||   ||| | ||| | | | |   | ||||||  |
Db         98 TGGCTGGGTTGGTGGATGGCTGGATGGCTGGGTTGGTGGATTGTTGGATTGCTGGTGACT 39

Qy        557 CAGTAGGGCTGCTTGCTT 574
                ||  | |  ||| ||
Db         38 GGTTATTGGTTGTTGGTT 21


RESULT 16
US-09-728-446-354/c
; Sequence 354, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
```

```
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 354
;   LENGTH: 436
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(436)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-354

  Query Match             7.5%;  Score 74.8;  DB 9;  Length 436;
  Best Local Similarity   61.1%;  Pred. No. 3.9e-08;
  Matches  121;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy        377 TTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 436
              ||||| |  ||   | | ||||| |||| | | ||| | |||  | ||||| | | ||
Db        226 TTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGC 167

Qy        437 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
              |||   ||| | ||| | | |||| | | ||| | | |||||  |||| | | ||| |
Db        166 TGGATGGCTGGGTTGGTGGATGGCTGGATGGCTGGATGGCTGGTTTGCTGGATGGCTGGA 107

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 556
              ||||  | ||| | | | |||||   ||| | ||| | | | |   | ||||||  |
Db        106 TGGCTGGGTTGGTGGATGGCTGGATGGCTGGGTTGGTGGATTGTTGGATTGCTGGTGACT 47

Qy        557 CAGTAGGGCTGCTTGCTT 574
                ||  | |  ||| ||
Db         46 GGTTATTGGTTGTTGGTT 29


RESULT 17
US-10-027-632-273001/c
; Sequence 273001, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 273001
;   LENGTH: 1006
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-273001

  Query Match             7.4%;  Score 73.8;  DB 13;  Length 1006;
  Best Local Similarity   45.9%;  Pred. No. 5.8e-08;
  Matches  295;  Conservative    0;  Mismatches  342;  Indels    6;  Gaps    1;

Qy        247 GCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 306
              | |||  |||  |  ||   | | | | | |||   | | | ||   |||  |||||
Db        924 GATGAATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGTAGGATG 865

Qy        307 GCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAG 366
              | | | ||   |||  |  |    | | | | | |||   | | | ||   |||  | |
Db        864 GATGGATGGATGGATGGATGATAGGATGGATGGATGGGTGGATGGATGGTAGGATGGATG 805

Qy        367 GGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 426
              |   | |||  | | |  |||||   | | | |||   | | | | | | | |||   | |
Db        804 GTAGGATGGATGGATGGTAGGATGGATGGATGGATGGATGGATGGATAGGTGGATAGGTG 745

Qy        427 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 486
              | | | | | |||   | | |   |||      |   | |   | | | | |||   | |
Db        744 GATAGATAGGTGGATGGATGGGCAGCTAAGGTGGTAGGTGAGTGGATGGATGGATGGATG 685

Qy        487 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 546
              | | | | | |||   | | | | | | | |     | | | | | | | |||   | |
Db        684 GATGGATGGATGGATGGATGGATAGGTAAGTGGATGGRTGGATGGATGGATAGGTA 625

Qy        547 GCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTT 606
                ||   ||   |   |   | | | |   | || ||| ||       |     ||  |
Db        624 AGTGGATGGGTGGATAGATGGATAGGTGTTTGAGTGGATGGATAAATAGATGTCTGAGTG 565

Qy        607 GATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 666
              |||   |  || |||| | | |  ||| |||||| | |   |     || |
Db        564 GATGAATGCATTGGTGGGTGGATGGGTAGGCAGCTGGGTAGATGTATGAAGGAGTGAG-- 507

Qy        667 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCT 726
              |  ||   | | ||      ||| |  | | |   ||    || || |   |
Db        506 ----GGAGTTTGGATAGGTATACGAAAGGCCAGCAGAGAGCCTTCGCTGGATGTTTCTCA 451

Qy        727 GCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCT 786
              |  |||||      || |      | || | ||  || ||   | | || ||   || |
Db        450 GTGAGTAGATGGATTGTGGGAGAGGGAGAGATGAATGGATGGATGGATGGATGGATGGAT 391

Qy        787 GCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTA 846
              |   | | || ||   || || ||| | || ||  ||   |  ||  || | |  ||
Db        390 GGATGGATGGATGGATGGATGGTAGGATGGATGGATGGATGGATGGATGATAGGATGGAT 331

Qy        847 GTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGA 889
              |   ||  |  ||  || ||  | || || |  |  |   |||
Db        330 GGATGGGTGGATGGATGGTAGGATGGATGGTAGGATGGATGGA 288


RESULT 18
US-10-027-632-272985/c
; Sequence 272985, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 272985
```

```
;   LENGTH: 572
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-272985

  Query Match             7.3%;  Score 73.4;  DB 13;  Length 572;
  Best Local Similarity   46.5%;  Pred. No. 7.7e-08;
  Matches  236;  Conservative    0;  Mismatches  271;  Indels    0;  Gaps    0;

Qy        331 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCT 390
              ||| |||   |     |||||| || | |  || |    ||||||| |||   |||
Db        568 GCTGGCTGTATTAAAGGCTTGCCTAGAGCATGGTTGTCAGGCTGGCTGGCTGTCAGGCTG 509

Qy        391 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 450
                  |     || |   ||| |       |||   || | | |||| |  |||
Db        508 AGCAAGGGTTTATCTGGAAGGCTGGTCAAGATGCAGTCTGGGTGGCTGCCTGGCTGAGAG 449

Qy        451 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 510
              ||| | || |   |   ||| |||    |||   ||  ||       ||||| |    | ||
Db        448 GCTGGTTGCCTGGGAAGCTGGCTGAGAGGCTGCCTGCCTGCAAGTCTGGCTTAGAGGGTT 389

Qy        511 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTT 570
               ||    |||  ||| | ||    ||| |||||   ||       |   || ||   |||
Db        388 ACTGAGAGGCTGGCTGGATTAGAGGCTTGCCTGGAGGCCAKCTAGTAGGTGGGCTGGCTG 329

Qy        571 GCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTA 630
                |||    |   |  | | | |      |||| | |   |||   ||||||  |||
Db        328 AGAGGCTAGCTGCTTGAAAGGGCAGTTGAGAGGCTGGCTGAGAGGCTGGCTGGGAGGCTG 269

Qy        631 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 690
              | | |  |||||  | ||   ||    ||| |  |||    ||      ||   |||
Db        268 TTTGAGAGGCTGGCTTCCTGAGAGACTGGCTCAGAGGCTGCCTGGTTGAGAGGCTGGCTT 209

Qy        691 GTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTA 750
               |||  |||||  |   || |  ||       |  |||||  ||     ||||   | |
Db        208 AGAGGCTTCCTGAGAGCCTGGCTGGTTGAGAAGCTTGCTAAGAGACTGGCTGGCAGGTTG 149

Qy        751 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 810
              | ·||      |  |||| |   |||  ||| ||  ||  |    |  |  ||   |||
Db        148 GCTGGCTGAGAGATCTGCCAAGGGGGTGGCTTGGAATCTTGCCCAGAGGTGGGCTGGCTC 89

Qy        811 GTAGGGCTGCCTGCTGGCTTGCTTGCT 837
                ||  |    | |  || |   |||
Db         88 CCTGGTTTAGAGGTTACCTCGAAGGCT 62


RESULT 19
US-09-728-446-295/c
; Sequence 295, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 295
;   LENGTH: 388
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(388)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-295

  Query Match             7.3%;  Score 73.2;  DB 9;  Length 388;
  Best Local Similarity   63.8%;  Pred. No. 9e-08;
  Matches  111;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        377 TTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 436
              ||||| |  ||   | | |||||  |||| | | ||| | ||||  | |||||  | | ||
Db        177 TTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGC 118

Qy        437 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
              |||   ||| | ||| | | ||||  | | ||| | | |||||  |||| | | ||| |
Db        117 TGGATGGCTGGGTTGGTGGATGGCTGGATGGCTGGATGGCTGGTTTGCTGGATGGCTGGA 58

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              ||||  | ||| | | | |||||   ||| | ||| | | | |   | ||||||
Db         57 TGGCTGGGTTGGTGGATGGCTGGATGGCTGGGTTGGTGGATTGTTGGATTGCTG 4


RESULT 20
US-09-922-217-1058/c
; Sequence 1058, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Jiang, Yuqiu
;  APPLICANT:  Smith, Carole Lynn
;  APPLICANT:  King, Gordon E.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
;  TITLE OF INVENTION:  OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C13
;  CURRENT APPLICATION NUMBER: US/09/922,217
;  CURRENT FILING DATE:  2001-08-03
;  NUMBER OF SEQ ID NOS: 1124
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1058
;   LENGTH: 15720
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-922-217-1058

  Query Match             7.3%;  Score 73.2;  DB 9;  Length 15720;
  Best Local Similarity   45.3%;  Pred. No. 5.6e-08;
  Matches  302;  Conservative    0;  Mismatches  363;  Indels    1;  Gaps    1;

Qy        282 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 341
              || ||| ||  |||   | ||   || ||| ||| |||   ||     || ||| | |
Db       5254 TGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGAAGGGGTGGTCATGGTGGTTGGTGGAG 5195

Qy        342 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 401
              |||  |       ||  ||     |||  | ||   || |  | ||   ||| |  |
Db       5194 GGCTGGGAGTGGTAGTTGAGGAAGGGGTGGTTGTGGTGGTTATTGGAGAGCTGGGGGTGG 5135

Qy        402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461
              || ||| |     | ||| |    || ||| | |  |||  |      || |   |||
Db       5134 TGGTTGGTGAAGGGGTGGTCATGGTGGTTGTTGGAGGGCTGGGAGTGGTGGTTATTGGAG 5075

Qy        462 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 521
               ||| |       | | |   |  | | | |     |  |  | |   ||| | |
Db       5074 AGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTATTGGAGGGCTGGGAGTGG 5015

Qy        522 TGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGG 581
              || ||| | |   |||||      || |  | |||  |  ||       || ||| |   |
Db       5014 TGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTGGTGGAG 4955

Qy        582 GGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTG-CTAGTAGGGCTGC 640
              ||  || |    ||  || |||   | | |||| |||| | ||  |  |  | ||  |
```

```
Db       4954 GGCTGGGAGTGGTGGTTGGTGGAAGGGTGGTAGTGCTGGTTGGTGGAGTGATGGGTGTAG 4895

Qy        641 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCC 700

Db       4894 TCGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTCGTTGTTGGAGGGC 4835

Qy        701 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 760

Db       4834 TGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGATTGTTGGAGGACTGGGAGTGGTGG 4775

Qy        761 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 820

Db       4774 TTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTGGTGGAGGGC 4715

Qy        821 CTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTG 880

Db       4714 TGGGAGTGGTGGTTGGTGGAAGGGTAGTAGTGCTGGTTGGTGGAGTGATGGGCGTAGTCG 4655

Qy        881 CAGCTGGGAGAACAGAGTAGGGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAG 940

Db       4654 TTGGAGGACTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTGTTGGAGGGCTGG 4595

Qy        941 ACGAGG 946

Db       4594 GAGTGG 4589


RESULT 21
US-09-833-263-1058/c
; Sequence 1058, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Meagher, Madeleine J.
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C12
;  CURRENT APPLICATION NUMBER: US/09/833,263
;  CURRENT FILING DATE:  2001-04-10
;  NUMBER OF SEQ ID NOS: 1093
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1058
;   LENGTH: 15720
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-833-263-1058

  Query Match             7.3%;  Score 73.2;  DB 10;  Length 15720;
  Best Local Similarity   45.3%;  Pred. No. 5.6e-08;
  Matches  302;  Conservative    0;  Mismatches  363;  Indels    1;  Gaps    1;

Qy        282 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 341

Db       5254 TGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGAAGGGGTGGTCATGGTGGTTGGTGGAG 5195

Qy        342 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 401

Db       5194 GGCTGGGAGTGGTAGTTGAGGAAGGGGTGGTTGTGGTGGTTATTGGAGAGCTGGGGGTGG 5135

Qy        402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461

Db       5134 TGGTTGGTGAAGGGGTGGTCATGGTGGTTGTTGGAGGGCTGGGAGTGGTGGTTATTGGAG 5075

Qy        462 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 521

Db       5074 AGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTATTGGAGGGCTGGGAGTGG 5015

Qy        522 TGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGG 581

Db       5014 TGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTGGTGGAG 4955
```

```
Qy        582 GGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTG-CTAGTAGGGCTGC 640

Db       4954 GGCTGGGAGTGGTGGTTGGTGGAAGGGTGGTAGTGCTGGTTGGTGGAGTGATGGGTGTAG 4895

Qy        641 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCC 700

Db       4894 TCGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTCGTTGTTGGAGGGC 4835

Qy        701 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 760

Db       4834 TGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGATTGTTGGAGGACTGGGAGTGGTGG 4775

Qy        761 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 820

Db       4774 TTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTGGTGGAGGGC 4715

Qy        821 CTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTG 880

Db       4714 TGGGAGTGGTGGTTGGTGGAAGGGTAGTAGTGCTGGTTGGTGGAGTGATGGGCGTAGTCG 4655

Qy        881 CAGCTGGGAGAACAGAGTAGGGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAG 940

Db       4654 TTGGAGGACTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTGTTGGAGGGCTGG 4595

Qy        941 ACGAGG 946

Db       4594 GAGTGG 4589


RESULT 22
US-10-025-380-1058/c
; Sequence 1058, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Jiang, Yuqiu
;  APPLICANT:  Smith, Carole L.
;  APPLICANT:  King, Gordon E.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  APPLICANT:  Skeiky, Yasir A. W.
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick Thomas S.
;  APPLICANT:  Carter, Darrick
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
;  TITLE OF INVENTION:  OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C14
;  CURRENT APPLICATION NUMBER: US/10/025,380
;  CURRENT FILING DATE:  2001-12-19
;  NUMBER OF SEQ ID NOS: 1129
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1058
;   LENGTH: 15720
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-025-380-1058

  Query Match             7.3%;  Score 73.2;  DB 13;  Length 15720;
  Best Local Similarity   45.3%;  Pred. No. 5.6e-08;
  Matches  302;  Conservative    0;  Mismatches  363;  Indels    1;  Gaps    1;

Qy        282 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCT 341

Db       5254 TGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGAAGGGGTGGTCATGGTGGTTGGTGGAG 5195

Qy        342 GGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCC 401
```

```
Db       5194 GGCTGGGAGTGGTAGTTGAGGAAGGGGTGGTTGTGGTGGTTATTGGAGAGCTGGGGGTGG 5135

Qy        402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461
              || ||| |     | ||| |    || ||| | |  |||  |       || |   |||
Db       5134 TGGTTGGTGAAGGGGTGGTCATGGTGGTTGTTGGAGGGCTGGGAGTGGTGGTTATTGGAG 5075

Qy        462 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 521
               ||| |      | | |   | | | |     |  |  | |   |||  |  |
Db       5074 AGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTATTGGAGGGCTGGGAGTGG 5015

Qy        522 TGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGG 581
              || ||| | |   |||||      || |  | |||  |  ||      || ||| |   |
Db       5014 TGGTTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTGGTGGAG 4955

Qy        582 GGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTG-CTAGTAGGGCTGC 640
              ||  || |    ||  || |||   | | |||| |||| | ||  |  |  | ||  |
Db       4954 GGCTGGGAGTGGTGGTTGGTGGAAGGGTGGTAGTGCTGGTTGGTGGAGTGATGGGTGTAG 4895

Qy        641 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCC 700
              | |   |   |  |||      || || | |  |||  | || | |  | || ||    |
Db       4894 TCGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTCGTTGTTGGAGGGC 4835

Qy        701 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 760
              ||||      |   ||  |  ||| | |     |  | |||    || | || | ||
Db       4834 TGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGATTGTTGGAGGACTGGGAGTGGTGG 4775

Qy        761 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 820
              | |   |   | |||      || || | |  |||  | || | || | |  ||   ||
Db       4774 TTGTTGGAGGGCTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTGGTGGAGGGC 4715

Qy        821 CTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTG 880
               |  |   || ||| | |   | ||||  | || | |||    |      ||||      |
Db       4714 TGGGAGTGGTGGTTGGTGGAAGGGTAGTAGTGCTGGTTGGTGGAGTGATGGGCGTAGTCG 4655

Qy        881 CAGCTGGGAGAACAGAGTAGGGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAG 940
               |  ||      || |  ||   |  |    |  |     | ||  ||  ||||||  |
Db       4654 TTGGAGGACTGGGAGTGGTGGTTGGTGGAGGGGTGGTTGTGGTGGTTGTTGGAGGGCTGG 4595

Qy        941 ACGAGG 946
                | ||
Db       4594 GAGTGG 4589
```

RESULT 23
US-10-017-161-2193
; Sequence 2193, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2193
; LENGTH: 1297
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(1297)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(1097)
US-10-017-161-2193

Query Match 7.3%; Score 73; DB 12; Length 1297;
Best Local Similarity 58.5%; Pred. No. 8.5e-08;
Matches 183; Conservative 0; Mismatches 125; Indels 5; Gaps 3;

```
Qy        220 GCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGG 279
              | | | | | |||| |||  | ||||||||          |    | || || ||||||
Db        644 GTTTCTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTTCTTGTTTTCTCGCTTT 703

Qy        280 CCTGCTTGCTGAGGGGACAGT-AGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTC 338
              | | ||| ||   |   |  |   || ||||| ||     ||  |    | ||||| ||
Db        704 CTTTCTTTCTTTTGTTTCTTTCCTGCTTGCTTTCTTGCTTGATTGCTTTCGTGCTTTCTT 763

Qy        339 GCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTG 398
              |||  | || || ||          |   | | ||  ||| || |      || |  ||
Db        764 GCTTTCTTGTTTTCT---TTCTTTCTTTTGTTTCTTTCTTTCTTGCTTCCTTGTTTTCTT 820

Qy        399 GCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGC-TCGCT 457
              || | |||||||||| |||  | ||||| ||||||  || || | ||| |||    |  ||
Db        821 GCTTTCTTGCTTGCTTGCTTTCGTGCTTTCTTGCTTTCTTGCTTTCTTTCTTTTGTTTCT 880

Qy        458 GGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCT 517
                | |||||||| || |||  | || || ||||||  || || |||||||| |  |||
Db        881 TTCTTGCTTGCTTTCTTGCTTCCTTGTTTTCTTGCTTTCTTGCTTGCTTGCTTTCGTGCT 940

Qy        518 GGCCTGCTTGCTT 530
                | ||||| |||
Db        941 TTCTTGCTTTCTT 953
```

RESULT 24
US-09-728-446-362/c
; Sequence 362, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
; TITLE OF INVENTION: and Mutant Cells and Mutant Animals Defined Thereby
; FILE REFERENCE: LEX-0101-USA
; CURRENT APPLICATION NUMBER: US/09/728,446
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/168,270
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 1461
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 362
; LENGTH: 409
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(409)
; OTHER INFORMATION: n = A,T,C or G
US-09-728-446-362

Query Match 7.2%; Score 72.2; DB 9; Length 409;
Best Local Similarity 63.2%; Pred. No. 1.5e-07;
Matches 110; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy        377 TTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 436
              |||||  |  ||   | | |||||  |||| | | ||| | ||||  | |||||  | | ||
Db        205 TTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGC 146

Qy        437 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
              |||   ||| | ||| | | ||||  | | ||| | | |||||  |||| | | ||| |
Db        145 TGGATGGCTGGGTTGGTGGATGGCTGGATGGCTGGATGGCTGGTTTGCTGGATGGCTGGA 86

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              ||||  | ||| | | | |||||   ||| | ||| | | | |   | ||| ||
Db         85 TGGCTGGGTTGGTGGATGGCTGGATGGCTGGGTTGGTGGATTGTTGGATTGNTG 32
```

RESULT 25

```
US-10-027-632-113494
; Sequence 113494, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 113494
;   LENGTH: 3213
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-113494

  Query Match             7.2%;  Score 71.6;  DB 13;  Length 3213;
  Best Local Similarity   47.0%;  Pred. No. 1.6e-07;
  Matches  287;  Conservative    0;  Mismatches  319;  Indels    4;  Gaps    2;

Qy        231 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 290
Db       1951 GGTGGATGGATGGATGAGTAGGTGGATGGATGGATGGATAGGTGGGTGGATGGGTGGGTG 2010

Qy        291 AGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTT 350
Db       2011 AGTGGATGGATGGATGGATGGATGGATGGATGGATGGATGGGTAGGTGAATGGGTC 2070

Qy        351 GCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTT 410
Db       2071 AGTGGATGGA-TGGATGGATGATGGATAGGTGAATGGGTCAGTGGATGGATGGATGGGTA 2129

Qy        411 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 470
Db       2130 GGTGGATGGATCTATGGGTAGGTGGATGGATCTATGGGTAGGTGGATGGGTGGGTGGGTA 2189

Qy        471 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 530
Db       2190 GATGGATGGCCAGATGGGTGGATGACTGGGTAGATGGATAGGTGAGTGGATGGATGGATG 2249

Qy        531 GCTCGCTGGCCTGCTTGCTGAGGGGACAGT---AGGGCTGCTTGCTTGCTAAGGGGACGG 587
Db       2250 GATGGATAGATGGATGGGTAGGTGGACGGTGGATGGATGAATGGGTAGGTGGATGGATGG 2309

Qy        588 TACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTG 647
Db       2310 ATAAGTTAATGAATGGATGGGTAGGTGGACGGATGGACGGGTAGGTGGATGGATGGATAG 2369

Qy        648 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTG 707
Db       2370 GTGGGTGGATGGATGGGTGAGTGGATGGGTGGGTGAGTGGGTGGATGGATGGATGGATGG 2429

Qy        708 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTG 767
Db       2430 ATGAATGGGTCGGTGGATGGACGGATGGGTAGGTGGATGGATCTATGGGTAGGTGGATGG 2489

Qy        768 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGG 827
Db       2490 GTGGGTGGGTGGATGGGTGGGTGGGTAGATGGATGGATGGATGGGTGGACGAGTGGGTAG 2549

Qy        828 CTTGCTTGCT 837
Db       2550 GTAGATAGGT 2559


RESULT 26
US-09-728-446-736/c
; Sequence 736, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 736
;   LENGTH: 392
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(392)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-736

  Query Match             7.1%;  Score 71.4;  DB 9;  Length 392;
  Best Local Similarity   62.4%;  Pred. No. 2.3e-07;
  Matches  111;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        397 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 456
Db        182 TTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGC 123

Qy        457 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 516
Db        122 TGGATGGCTGGGTTGGTGGATGGCTGGATGGCTGGATGGCTGGTTTNCTGGATGGCTGGA 63

Qy        517 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTT 574
Db         62 TGGCTGGGTTGGTGGATGGCTGGATGGCTGGGTTGGTGGATTGTTGGATAGCTGGTTT 5


RESULT 27
US-09-728-446-246/c
; Sequence 246, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 246
;   LENGTH: 379
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
```

```
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(379)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-246

  Query Match             7.1%;  Score 70.8;  DB 9;  Length 379;
  Best Local Similarity   59.1%;  Pred. No. 3.1e-07;
  Matches  117;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        377 TTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 436
              ||||| |  ||   | | |||||  |||| | | ||| | ||||  | |||||| | | ||
Db        198 TTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGC 139

Qy        437 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
              |||   ||| |  || |   || |    | ||| | | |||||  |||| | | ||| |
Db        138 TGGATGGCTGGGNTGGTGNATGNCTGNATGGCTGGATGGCTGGTTTGCTGGATGGCTGGA 79

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 556
              ||||  | ||| | | | |||||   ||| | ||| | | | |   | ||||||  |
Db         78 TGGCTGGGTTGGTGGATGGCTGGATGGCTGGGTTGGTGGATTGTTAGATTGCTGGTGACT 19

Qy        557 CAGTAGGGCTGCTTGCTT 574
              ||  | |  ||| ||
Db         18 GGTTATTGGTTGTTGGTT 1


RESULT 28
US-09-728-446-250/c
; Sequence 250, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 250
;   LENGTH: 427
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(427)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-250

  Query Match             7.1%;  Score 70.8;  DB 9;  Length 427;
  Best Local Similarity   59.4%;  Pred. No. 3e-07;
  Matches  117;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy        377 TTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 436
              ||||| |  ||   | |  ||||  |||| | | ||| | ||||  | |||||| | | ||
Db        217 TTGCTGGATGGCTGGATGNCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGC 158

Qy        437 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
              |||   ||| | ||| |   ||||  | | ||| | | |||||  ||||   | ||| |
Db        157 TGGATGGCTGGGTTGGTGNATGGCTGGATGGCTGGATGGCTGGTTTGCTGNATGGCTGGA 98

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 556
              ||||  | ||| | | | |||||   ||| | ||| | | | |   | ||||||  |
Db         97 TGGCTGGGTTGGTGGATGGCTGGATGGCTGGGTTGGTGGATTGTTGGATTGCTGGTGACT 38

Qy        557 CAGTAGGGCTGCTTGCT 573
              ||  | |  ||| |
Db         37 GGTTATTGGTTGTTGGT 21


RESULT 29
US-09-728-446-325/c
; Sequence 325, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 325
;   LENGTH: 410
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(410)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-325

  Query Match             7.1%;  Score 70.6;  DB 9;  Length 410;
  Best Local Similarity   62.6%;  Pred. No. 3.4e-07;
  Matches  109;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy        377 TTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 436
              |||||  |  ||   | | |||||  |||| | | ||| | ||||  | ||||| | | ||
Db        200 TTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGC 141

Qy        437 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 496
              |||   ||| | ||| | | ||||  | | ||| | | |||||  |||| | | ||| |
Db        140 TGGATGGCTGGGTTGGTGGATGGCTGGATGGCTGGATGGCTGGTTTGCTGGATGGCTGGA 81

Qy        497 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              ||||  | ||  | | | |||||   ||| | ||| | | | |   | | ||||
Db         80 TGGCTGGGTTCGTGGATGGCTGGATGGCTGGGTTGGTGGATTGTTGGATNGCTG 27


RESULT 30
US-09-764-891-8904
; Sequence 8904, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC006
;  CURRENT APPLICATION NUMBER: US/09/764,891
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 10231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8904
;   LENGTH: 368
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-891-8904

  Query Match             6.9%;  Score 69.4;  DB 11;  Length 368;
  Best Local Similarity   60.3%;  Pred. No. 6.4e-07;
  Matches  132;  Conservative    0;  Mismatches   86;  Indels    1;  Gaps    1;

Qy        325 AGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGT 384
              | || ||||| || ||| || |||||||||   | |       |  | |  | || || |
Db         99 ATGCTTGCTTTCTTGCTTGCTTGCTTGCTTTCGTGCTTTCTTGCTTTCCCGTTTTCTTGC 158

Qy        385 AGGGCTGCTCGC-TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 443
```

```
Db        159 TTTCTTTCTTTCTTTTGTTTCTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTT 218

Qy        444 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 503
              |||| || ||| |    |  ||  ||  |  ||  |||||||||| || ||| || ||
Db        219 CTTGTTTTCTCCTTTCTTTCTTTCTTTTGTTTCTTTCCTGCTTGCTTTCTTGCTTGCTTG 278

Qy        504 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 542
              ||| | ||||  || || | |||| || ||  ||  | |
Db        279 CTTTCGTGCTTTCTTGCTTTCTTGTTTTCTTTCTTTCTT 317


RESULT 31
US-09-764-891-10047
; Sequence 10047, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC006
;  CURRENT APPLICATION NUMBER: US/09/764,891
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 10231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10047
;   LENGTH: 368
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-891-10047

  Query Match             6.9%;  Score 69.4;  DB 11;  Length 368;
  Best Local Similarity   60.3%;  Pred. No. 6.4e-07;
  Matches  132;  Conservative    0;  Mismatches   86;  Indels    1;  Gaps    1;

Qy        325 AGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGT 384
              | || |||||| || ||| || ||||||||   | |       |  | |  | || || |
Db         99 ATGCTTGCTTTCTTGCTTGCTTGCTTGCTTTCGTGCTTTCTTGCTTTCCCGTTTTCTTGC 158

Qy        385 AGGGCTGCTCGC-TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 443
              | ||  | |    | ||| |||||| |||  | ||||||||||||  |  || |
Db        159 TTTCTTTCTTTCTTTTGTTTCTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTCGTGCTTT 218

Qy        444 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 503
              |||| || |||  |    |  ||  ||  |  ||  |||||||||| || ||| || ||
Db        219 CTTGTTTTCTCCTTTCTTTCTTTCTTTTGTTTCTTTCCTGCTTGCTTTCTTGCTTGCTTG 278

Qy        504 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 542
              ||| | ||||  || || | |||| || ||  ||  | |
Db        279 CTTTCGTGCTTTCTTGCTTTCTTGTTTTCTTTCTTTCTT 317


RESULT 32
US-09-728-446-465
; Sequence 465, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 465
;   LENGTH: 397
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(397)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-465

  Query Match             6.7%;  Score 67;  DB 9;  Length 397;
  Best Local Similarity   65.5%;  Pred. No. 2.2e-06;
  Matches   97;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy        392 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 451
              ||| |||   | | || |  | |  |||  |||| | ||||| | | ||  | ||||| |
Db         32 CTCCCTGTTGTTCCTGTTGCCACTTTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGG 91

Qy        452 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 511
              | |||||  |||| | ||||| | | ||  | ||||| | | |||||  |||| | |||
Db         92 ATTGCTGGATTGCTGGATTGCTGGATTGCTGGATTGCTGGATNGCTGGGTTGCTGGGTTG 151

Qy        512 CTCGCTGGCCTGCTTGCTTGCTCGCTGG 539
              || | | ||  | |||||  | | |||||
Db        152 CTGGATTGCTGGATTGCTGGATTGCTGG 179


RESULT 33
US-10-040-739-1102/c
; Sequence 1102, Application US/10040739
; Publication No. US20020173635A1
;  GENERAL INFORMATION:
;       APPLICANT: Jacobs, Kenneth
;                  McCoy, John
;                  LaVallie, Edward
;                  Racie, Lisa
;                  Merberg, David
;                  Treacy, Maurice
;                  Spaulding, Vikki
;       TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
;       NUMBER OF SEQUENCES: 1519
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Genetics Institute, Inc.
;            STREET: 87 CambridgePark Drive
;            CITY: Cambridge
;            STATE: Massachusetts
;            COUNTRY: U.S.A
;            ZIP: 02140
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy Disk
;            COMPUTER: IBM PC Compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/040,739
;            FILING DATE: 07-Jan-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/036,520
;            FILING DATE: 03-JUN-1998
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Brown, Scott A.
;            REGISTRATION NUMBER: 32,724
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (617) 498-8224
;            TELEFAX: (617) 876-5851
;  INFORMATION FOR SEQ ID NO: 1102:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 368 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1102:
US-10-040-739-1102

  Query Match             6.5%;  Score 65.2;  DB 13;  Length 368;
```

```
  Best Local Similarity   68.5%;  Pred. No. 5.5e-06;
  Matches  102;  Conservative    0;  Mismatches   46;  Indels    1;  Gaps    1;

Qy        402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461
              | ||| |||||| ||| | |||||||| ||| |||  | ||||| ||||||  |  |
Db        231 TTCTTTCTTGCTTGCTTTCTTGCTTGCTNGCTTGCTTTCGTGCTTTCTTGCTTTCCTGTT 172

Qy        462 TGCTTGCTTGCTCGCTGGCCTGC-TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 520
              | ||| ||| ||  ||  | |   || ||| || ||| || | ||||||||||| |||  |
Db        171 TNCTTTCTTTCTTTCTTTCTTTTGTTTCTTTCTTGCTTGCTTTCTTGCTTGCTTGCTTTC 112

Qy        521 CTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
               ||||| |||| |  || || | ||| ||
Db        111 GTGCTTTCTTGTTTTCTCGCNTTCTTTCT 83


RESULT 34
US-09-874-923-7/c
; Sequence 7, Application US/09874923
; Patent No. US20020081320A1
; GENERAL INFORMATION:
;  APPLICANT: Reed, Steven G.
;  APPLICANT:  Campos-Neto, Antonio
;  APPLICANT:  Webb, John R.
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Bhatia, Ajay
;  APPLICANT:  Coler, Rhea
;  APPLICANT:  Probst, Peter
;  APPLICANT:  Brannon, Mark
;  TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
;  TITLE OF INVENTION:  THERAPY AND DIAGNOSIS OF LEISHMANIASIS
;  FILE REFERENCE: 210121.420C8
;  CURRENT APPLICATION NUMBER: US/09/874,923
;  CURRENT FILING DATE:  2001-06-04
;  NUMBER OF SEQ ID NOS: 122
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 1771
;   TYPE: DNA
;   ORGANISM: Leishmania tropica
US-09-874-923-7

  Query Match             6.4%;  Score 64.2;  DB 9;  Length 1771;
  Best Local Similarity   48.7%;  Pred. No. 7.6e-06;
  Matches  320;  Conservative    0;  Mismatches  328;  Indels    9;  Gaps    5;

Qy        207 GACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 266
              | | ||  |   | |||     |  ||  ||| |||  | |||         |||  || |
Db        881 GCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGTGCATGACATCCAGGCGGGCC 822

Qy        267 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 326
              || | ||||   |||| |  |||  | |   ||  |  |||| | ||   ||| | | |
Db        821 GCCTCCTCGGCCGCCT-CCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGAC 763

Qy        327 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 386
              ||  || ||||||    | ||| ||       | | |   | || || |   ||    ||
Db        762 GCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAG 703

Qy        387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 446
               | || ||  | |  || ||| | |||  ||||||   ||  ||  ||| | |  | |
Db        702 CTCCGCGCGGAGACGCGC-TGCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGT 644

Qy        447 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 506
              || ||    | | |     || | |||  ||||||    |  ||  ||| | |  | |
Db        643 GCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGGCCTCGT 584

Qy        507 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT---GCTTGCTGAGGGGACAGTAGG 563
              | ||  ||    | | | |   || | ||||   ||||   |||    | || |||     |
Db        583 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTG 524

Qy        564 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 623
               || || |   ||   || |  ||   ||  | |     || ||       ||| |   |

Db        523 CCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGCATGGCCTCCAGGCGGGCCG 464

Qy        624 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 683
               || || |      || | |  | ||  | ||    ||| | || || |  || |||   |
Db        463 CCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-GCCTCCTCGAGGGCCTGGACG 405

Qy        684 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 743
                 ||  |   |||  |  |  ||  |    ||| |    | || || |     |||| |
Db        404 CGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCGCCTCCTCGGCTTCCTCCAGC 345

Qy        744 GCTGC---TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 800
               | ||    ||  | |||||       |   |   ||  || || ||||| |      |||
Db        344 TCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGC 285

Qy        801 TGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 857
               |||  | ||  | || |||| || |   || | |      ||  |  |   ||||
Db        284 ATGGCCTCCAGGCGCGCCGCCTCCTCGGCAGCCTCCAGCTCCGCGCGGAGACGCGCT 228


RESULT 35
US-09-991-496-7/c
; Sequence 7, Application US/09991496
; Patent No. US20020169285A1
; GENERAL INFORMATION:
;  APPLICANT: Reed, Steven G.
;  APPLICANT:  Campos-Neto, Antonio
;  APPLICANT:  Webb, John R.
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Bhatia, Ajay
;  APPLICANT:  Coler, Rhea
;  APPLICANT:  Probst, Peter
;  APPLICANT:  Brannon, Mark
;  TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
;  TITLE OF INVENTION:  THERAPY AND DIAGNOSIS OF LEISHMANIASIS
;  FILE REFERENCE: 210121.420C9
;  CURRENT APPLICATION NUMBER: US/09/991,496
;  CURRENT FILING DATE:  2001-11-20
;  NUMBER OF SEQ ID NOS: 137
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 1771
;   TYPE: DNA
;   ORGANISM: Leishmania tropica
US-09-991-496-7

  Query Match             6.4%;  Score 64.2;  DB 10;  Length 1771;
  Best Local Similarity   48.7%;  Pred. No. 7.6e-06;
  Matches  320;  Conservative    0;  Mismatches  328;  Indels    9;  Gaps    5;

Qy        207 GACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCT 266
              | | ||  |   | |||     |  ||  ||| |||  | |||         |||  || |
Db        881 GCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGCCCTCGTGCATGACATCCAGGCGGGCC 822

Qy        267 GCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAG 326
              || | ||||   |||| |  |||  | |   ||  |  |||| | ||   ||| | | |
Db        821 GCCTCCTCGGCCGCCT-CCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGAC 763

Qy        327 GCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAG 386
              ||  || ||||||    | ||| ||       | | |   | || || |   ||    ||
Db        762 GCGGGCCTGCTCGGCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAG 703

Qy        387 GGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 446
               | || ||  | |  || ||| | |||  ||||||   ||  ||  ||| | |  | |
Db        702 CTCCGCGCGGAGACGCGC-TGCCTCCTCGAGGGCCTGGACGCGGGCCTGCTCGGCCTCGT 644

Qy        447 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 506
              || ||    | | |     || | |||  ||||||    |  ||  ||| | |  | |
Db        643 GCATGACATCCAGGCGCGCCGCCTCCTCGAGGGCCTGGACACGGGCCTGCTCGGCCTCGT 584

Qy        507 GCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT---GCTTGCTGAGGGGACAGTAGG 563
              | ||  ||    | | | |   || | ||||   ||||   |||    | || |||     |
Db        583 GCATGACATCCAGGCGGGCCGCCTCCTCGGCCGCCTCCAGCTCCGCGCGGAGACGCGCTG 524
```

```
Qy        564 GCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGG 623
              || || |   ||   || |  ||   ||  | |     || ||       ||| |   |
Db        523 CCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGCATGGCCTCCAGGCGGGCCG 464

Qy        624 GCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGG 683
              || || |     || | |  | ||  | ||    ||| | || || |  || |||   |
Db        463 CCTCCTCGGCCTCCTCCAGCTCCGCGCGGAGACGCGCT-GCCTCCTCGAGGGCCTGGACG 405

Qy        684 GCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGG 743
              ||  |   |||  |  |  ||  |    ||| |  | || || |     |||| |
Db        404 CGGGCCTGCTCGGCCGCATGCATGACATCCAGGCGGGCCGCCTCCTCGGCTTCCTCCAGC 345

Qy        744 GCTGC---TAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 800
              | ||    ||  | |||||      |   |   ||  || || ||||| |      |||
Db        344 TCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGGAGCGGGCCTGCTCGGCCTCGTGC 285

Qy        801 TGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCT 857
              |||  | ||  | || |||| || |   || | |       ||  |  |   ||||
Db        284 ATGGCCTCCAGGCGCGCCGCCTCCTCGGCAGCCTCCAGCTCCGCGCGGAGACGCGCT 228


RESULT 36
US-10-027-632-141458/c
; Sequence 141458, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 141458
;   LENGTH: 681
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-141458

  Query Match             6.4%;  Score 63.8;  DB 13;  Length 681;
  Best Local Similarity   58.2%;  Pred. No. 1e-05;
  Matches  110;  Conservative    1;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        361 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 420
              ||  | |||||  || |||   || ||| ||    ||  | | ||| ||| ||  ||  |
Db        590 CACCATGGCTGGAGGATTGAGGGTGGGGATGGGATCTTTCTTTCTTTCTTTCTTTCTTTC 531

Qy        421 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 480
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||| ||  |
Db        530 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTCTC 471

Qy        481 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 540
              ||  ||: ||| ||  |   ||| ||| ||| ||  |   ||  ||| ||  ||  |
Db        470 TCTCTCTCTYTCTCTCTCTCTCTTTCTTTCTCTCTCTCTCTCTCTTCCTTCCTTCC 411

Qy        541 CTGCTTGCT 549
              | ||| ||
Db        410 TTCCTTCCT 402
```

```
RESULT 37
US-10-017-161-1581
; Sequence 1581, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1581
;   LENGTH: 13830
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(13830)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(486)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (4460)..(4952)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (5748)..(5918)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (6811)..(6954)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (10095)..(10238)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (10541)..(10707)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (13494)..(13630)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (2697)..(2796)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (13826)..(13830)
;   OTHER INFORMATION: a, t, c, g, unknown or other
US-10-017-161-1581

  Query Match             6.4%;  Score 63.6;  DB 12;  Length 13830;
  Best Local Similarity   52.7%;  Pred. No. 7.9e-06;
  Matches  138;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy         286 TGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCC 345
               |||| |||   | |   ||||||   |        | ||   ||| | |||  ||   ||
Db        4639 TGCTCAGGTACCTGCCCGCCTGCCCACCTGCCCACCTGTCCACCTACCTGCCTGCCCACC 4698

Qy         346 TGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCT 405
               |||   | |        ||    ||||   | |||      | |  |  |    ||| |
Db        4699 TGCCCACCGGCCTACTTGTCCACCTGCATACCTGCCCACCTGCCCACCTGTCCACCTACC 4758

Qy         406 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 465
               || ||||| || |||||| ||| | |  |    ||||| | | ||| |||   ||  |
Db        4759 CGCCTGCTCACTTGCCTGCCTGCCTACCTGTCCACCTGCATACCTGCCCGCCTACCCACC 4818
```

```
Qy        466 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 525
              || |  |   |  ||||||   | || || |   |||||  | | |||  ||  ||||||
Db       4819 TGTTCACCTACCTGCCTGCCCACCTGTTCACCTACCTGCCTCCCTGCCGGCCTGCCTGCC 4878

Qy        526 TGCTTGCTCGCTGGCCTGCTTG 547
              | | ||   ||| | |||  ||
Db       4879 TACCTGTCTGCTTGTCTGTCTG 4900


RESULT 38
US-09-764-872-812
; Sequence 812, Application US/09764872
; Publication No. US20030050231A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA125
;  CURRENT APPLICATION NUMBER: US/09/764,872
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 957
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 812
;   LENGTH: 31718
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-872-812

  Query Match             6.3%;  Score 62.8;  DB 11;  Length 31718;
  Best Local Similarity   69.7%;  Pred. No. 1.1e-05;
  Matches   85;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy        396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455
              ||  | | ||||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy        456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy        516 CT 517
              ||
Db      15036 CT 15037


RESULT 39
US-09-764-872-813
; Sequence 813, Application US/09764872
; Publication No. US20030050231A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA125
;  CURRENT APPLICATION NUMBER: US/09/764,872
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 957
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 813
;   LENGTH: 31718
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-872-813

  Query Match             6.3%;  Score 62.8;  DB 11;  Length 31718;
  Best Local Similarity   69.7%;  Pred. No. 1.1e-05;
  Matches   85;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy        396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455
              ||  | | ||||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy        456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy        516 CT 517
              ||
Db      15036 CT 15037


RESULT 40
US-09-764-891-9103
; Sequence 9103, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC006
;  CURRENT APPLICATION NUMBER: US/09/764,891
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 10231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9103
;   LENGTH: 31718
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-891-9103

  Query Match             6.3%;  Score 62.8;  DB 11;  Length 31718;
  Best Local Similarity   69.7%;  Pred. No. 1.1e-05;
  Matches   85;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy        396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455
              ||  | | ||||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy        456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
Db      14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy        516 CT 517
              ||
Db      15036 CT 15037


RESULT 41
US-09-764-891-9104
; Sequence 9104, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC006
;  CURRENT APPLICATION NUMBER: US/09/764,891
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 10231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9104
;   LENGTH: 31718
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-891-9104

  Query Match             6.3%;  Score 62.8;  DB 11;  Length 31718;
  Best Local Similarity   69.7%;  Pred. No. 1.1e-05;
  Matches   85;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy        396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455
              ||  | | ||||||||||| ||| || |||||||||||| ||| || ||||||||||||
Db      14916 CTTTCTTTCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTT 14975

Qy        456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515
              ||  | | ||| ||| ||  ||  | | ||| ||| |||  |    | ||| ||| ||
```

```
Db      14976 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTTTCTTTTTCTTTCTTTCTTT 15035

Qy        516 CT 517
              ||
Db      15036 CT 15037


RESULT 42
US-10-027-632-141457/c
; Sequence 141457, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 141457
;   LENGTH: 681
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-141457

  Query Match             6.3%;  Score 62.6;  DB 13;  Length 681;
  Best Local Similarity   58.2%;  Pred. No. 1.9e-05;
  Matches  110;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy        361 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 420
              ||  | |||||  || |||   || ||| ||    ||  | | ||| |||-||  ||  |
Db        590 CACCATGGCTGGAGGATTGAGGGTGGGGATGGGATCTTTCTTTCTTTCTTTCTTTCTTTC 531

Qy        421 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 480
                | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||| ||  |
Db        530 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTCTCTC 471

Qy        481 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 540
                ||  ||  ||| ||  |   ||| ||| ||| ||  |   ||  ||| ||  ||  |
Db        470 TCTCTCTCTCTCTCTCTCTCTCTTTCTTTCTCTCTCTCTCTCTCTTCCTTCCTTCC 411

Qy        541 CTGCTTGCT 549
                | ||| ||
Db        410 TTCCTTCCT 402


RESULT 43
US-10-027-632-141460/c
; Sequence 141460, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 141460
;   LENGTH: 681
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-141460

  Query Match             6.3%;  Score 62.6;  DB 13;  Length 681;
  Best Local Similarity   58.2%;  Pred. No. 1.9e-05;
  Matches  110;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy        361 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 420
              ||  | |||||  || |||   || ||| ||    ||  | | ||| |||  ||  |
Db        590 CACCATGGCTGGAGGATTGAGGGTGGGGATGGGATCTTTCTTTCTTTCTTTCTTTCTTTC 531

Qy        421 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 480
                | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||| ||  |
Db        530 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTYTCTTTCTTTCTTTCTCTCTCTC 471

Qy        481 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 540
                ||  ||  ||| ||  |   ||| ||| ||| ||  |   ||  ||| ||  ||  |
Db        470 TCTCTCTCTCTCTCTCTCTCTCTTTCTTTCTCTCTCTCTCTCTCTTCCTTCCTTCC 411

Qy        541 CTGCTTGCT 549
                | ||| ||
Db        410 TTCCTTCCT 402


RESULT 44
US-10-027-632-15974
; Sequence 15974, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15974
;   LENGTH: 726
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-15974

  Query Match             6.3%;  Score 62.6;  DB 13;  Length 726;
```

```
  Best Local Similarity   47.8%;  Pred. No. 1.9e-05;
  Matches  211;  Conservative    1;  Mismatches  225;  Indels    4;  Gaps    1;

Qy        386 GGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 445
              ||  | |  |  |  |||| |  ||  |      |||| |  |||  || |  ||||
Db          7 GGTGTCCCTGAGGCTCTGCAGGTATGGACATGCCTCTGCTGGTATGCATGCAGCTCTGCA 66

Qy        446 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 505
              |   | | | |  |||| |  |  |  |  |||| |  |||  ||  |  ||||
Db         67 GGTGGGGACACGGCTCTGCAGGTGGGGACATGGCTCTGCAGGTATGCACGTGGCTCTGCA 126

Qy        506 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGC 565
              |  | | | |  |||| |  |  |  |      |  |||||    |  | ||  |
Db        127 GGTGGGGACACGGCTCTGCAGGTGTCGACATGCCTC----TGCTGGTATGCACGCAGCTC 182

Qy        566 TGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGC 625
              |||  |   :  | ||  | | | |  ||||||  |   ||   |||  |  ||  |  |
Db        183 TGCAGGTGGRGACATGGCTCTGCAGGTATGCCTGCGGCTCTGCAGGTATGCATGTGGCTC 242

Qy        626 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGC 685
              |||  |||  |  |||  |  ||||  ||  |||        ||  | ||||  |  |  |
Db        243 TGCAGGTATGCATGCAGCTCTGCAGGTGGGGACATGCCTCTACCAGTATGCACGTGGCTC 302

Qy        686 TGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGC 745
              |||  ||  |  ||  |  ||||  |||  ||  |  |  ||||  ||  ||     |  |
Db        303 TGCAGGTGTGCACCCGGCTCTGCAGGTATGGACACAGCTCTGCAGGTGTGGACAGGGCTC 362

Qy        746 TGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGC 805
              |||  ||  |  |||  |  ||  |  ||  ||  |||    ||||  |||  |     |  |
Db        363 TGCAGGTGTGCATGCAGCTCTACAGGTGTGGATGCACCTCTGCAGGTATGCACATGGCCC 422

Qy        806 TGCTAGTAGGGCTGCCTGCTG 826
                ||  |   | |||||||
Db        423 CTGGAGCTGTCTTCCCTGCTG 443


RESULT 45
US-10-027-632-141459/c
; Sequence 141459, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 141459
;   LENGTH: 681
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-141459

  Query Match             6.2%;  Score 62.2;  DB 13;  Length 681;
  Best Local Similarity   57.7%;  Pred. No. 2.4e-05;
  Matches  109;  Conservative    1;  Mismatches   79;  Indels    0;  Gaps    0;

Qy        361 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 420
              ||  | |||||  || |||   || ||| ||    ||  | | ||| ||| ||  ||  |
Db        590 CACCATGGCTGGAGGATTGAGGGTGGGGATGGGATCTTTCTTTCTTTCTTTCTTTCTTTC 531

Qy        421 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 480
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||: ||| ||  |
Db        530 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTYTCTCTCTC 471

Qy        481 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 540
                ||  ||  ||| ||  |   ||| ||| ||| ||  |   ||  ||| ||  ||  |
Db        470 TCTCTCTCTCTCTCTCTCTCTCTTTCTTTCTCTCTCTCTCTCTCTTCCTTCCTTCC 411

Qy        541 CTGCTTGCT 549
                | ||| ||
Db        410 TTCCTTCCT 402


RESULT 46
US-10-027-632-174763/c
; Sequence 174763, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 174763
;   LENGTH: 2940917
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(2940917)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-174763

  Query Match             6.2%;  Score 62.2;  DB 13;  Length 2940917;
  Best Local Similarity   76.8%;  Pred. No. 8.1e-06;
  Matches   76;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy        402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461
              ||||||||||||| ||| || |||||||||||| ||| || |||||||||||| ||| |||
Db    2504045 TGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCC 2503986

Qy        462 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 500
                | | | | ||| | |   ||| ||| ||||||| ||
Db    2503985 ACTTCGTTCGTTCGTTCGTTCGCTCGCTCGCTCGCTCGC 2503947


RESULT 47
US-10-027-632-161036
; Sequence 161036, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
```

```
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE:  2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 161036
;   LENGTH: 863
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-161036

  Query Match             6.1%;  Score 61.4;  DB 13;  Length 863;
  Best Local Similarity   68.3%;  Pred. No. 3.5e-05;
  Matches  114;  Conservative    0;  Mismatches   51;  Indels    2;  Gaps    2;

Qy        352 CTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTG 411
              |||| ||    |  ||   |||||||||||||| | |||||||||||| | ||||| | |
Db        642 CTGATGGCTTGGCTGGCTGGCTGGCTTGCTAGCATGGCTGCTCGCTTGGCTGCTGGGTGC 701

Qy        412 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 471
              | | ||||  ||| |||   |  |||||  ||| |||   | |||||| ||||  || |
Db        702 TTGGGTGGCTGGCTGGCTGCTTTCCTGGCTGGCTGGCTGCATGGCTGGCTTGCTGACTAG 761

Qy        472 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 518
              | ||| | | ||||||||||| | |||| |||   |||||| | ||
Db        762 GT-GCTTGACGGCTTGCTTGCT-GGTGGCTTGCCGCCTTGCTGGTTG 806


RESULT 48
US-10-027-632-161037
; Sequence 161037, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 161037
;   LENGTH: 863
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-161037

  Query Match             6.1%;  Score 61.4;  DB 13;  Length 863;
  Best Local Similarity   68.3%;  Pred. No. 3.5e-05;
  Matches  114;  Conservative    0;  Mismatches   51;  Indels    2;  Gaps    2;

Qy        352 CTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTG 411
              |||| ||    |  ||   |||||||||||||| | |||||||||||| | ||||| | |
Db        642 CTGATGGCTTGGCTGGCTGGCTGGCTTGCTAGCATGGCTGCTCGCTTGGCTGCTGGGTGC 701

Qy        412 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 471
              | | ||||  ||| |||   |  |||||  ||| |||   | |||||| ||||  || |
Db        702 TTGGGTGGCTGGCTGGCTGCTTTCCTGGCTGGCTGGCTGCATGGCTGGCTTGCTGACTAG 761

Qy        472 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 518
              | ||| | | ||||||||||| | |||| |||   |||||| | ||
Db        762 GT-GCTTGACGGCTTGCTTGCT-GGTGGCTTGCCGCCTTGCTGGTTG 806


RESULT 49
US-10-027-632-161038
; Sequence 161038, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 161038
;   LENGTH: 863
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-161038

  Query Match             6.1%;  Score 61.4;  DB 13;  Length 863;
  Best Local Similarity   68.3%;  Pred. No. 3.5e-05;
  Matches  114;  Conservative    0;  Mismatches   51;  Indels    2;  Gaps    2;

Qy        352 CTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTG 411
              |||| ||    |  ||   |||||||||||||| | |||||||||||| | ||||| | |
Db        642 CTGATGGCTTGGCTGGCTGGCTGGCTTGCTAGCATGGCTGCTCGCTTGGCTGCTGGGTGC 701

Qy        412 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 471
              | | ||||  ||| |||   |  |||||  ||| |||   | |||||| ||||  || |
Db        702 TTGGGTGGCTGGCTGGCTGCTTTCCTGGCTGGCTGGCTGCATGGCTGGCTTGCTGACTAG 761

Qy        472 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTG 518
              | ||| | | ||||||||||| | |||| |||   |||||| | ||
Db        762 GT-GCTTGACGGCTTGCTTGCT-GGTGGCTTGCCGCCTTGCTGGTTG 806


RESULT 50
US-09-764-877-3185
; Sequence 3185, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
```

```
;  FILE REFERENCE: PC005
;  CURRENT APPLICATION NUMBER: US/09/764,877
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 4031
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3185
;   LENGTH: 223
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-877-3185

  Query Match             6.1%;  Score 61;  DB 10;  Length 223;
  Best Local Similarity   58.6%;  Pred. No. 5.1e-05;
  Matches  106;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy        369 CTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 428
              ||  | |||   | |   |  |  | |||    ||  |||||| | |||    |   |||
Db         43 CTCTTAGCTCAGTTGGGTGCTTAGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTAGGTGC 102

Qy        429 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 488
              ||| | |||    ||  |||||| | |||    || ||| | |||   |    ||||||
Db        103 TTGGTAGCTCAGTTGGGTGCTTGGTAGCTCAGGTGTTTGGTAGCTCAGTTAGGTGCTTGG 162

Qy        489 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 548
              | ||||  | |  ||||||  | ||||  | |  ||||||  | ||||  | |  ||||||
Db        163 TAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGA 222

Qy        549 T 549
              |
Db        223 T 223


RESULT 51
US-10-027-632-273402
; Sequence 273402, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 273402
;   LENGTH: 495
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-273402

  Query Match             6.1%;  Score 61;  DB 13;  Length 495;
  Best Local Similarity   57.9%;  Pred. No. 4.6e-05;
  Matches  106;  Conservative    1;  Mismatches   76;  Indels    0;  Gaps    0;

Qy        370 TGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 429
              || | |  || |||   ||  | | | | |  || ||| ||| | | | |  || | | |
Db        285 TGTTTGGATGGTAGGTTGGTGGATTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGGT 344

Qy        430 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 489
              || | | |||  || ||| ||| | | |||  || ||| ||| | | | |  || ||| |
Db        345 TGGTTGGTGGGTTGGTTGGTTGGTTGGTGGGTTGGTTGGTTGGTTGGTTGGTTGGT 404

Qy        490 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
                | | | |  ||:||| ||| | | | |  || | | ||| | | |    || ||| |
Db        405 GAGTTGGTTGGTTGSTTGGTTGGTTGGTTGGTGGGTTGGTGGGTTTGTTGGTTGGT 464

Qy        550 GAG 552
              |||
Db        465 GAG 467


RESULT 52
US-10-311-455-300
; Sequence 300, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Det
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 300
;   LENGTH: 6109
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 1763, 3313..3314, 5896
;   OTHER INFORMATION: n is a or g or c or t
US-10-311-455-300

  Query Match             6.1%;  Score 60.6;  DB 12;  Length 6109;
  Best Local Similarity   45.1%;  Pred. No. 4.1e-05;
  Matches  264;  Conservative    0;  Mismatches  319;  Indels    2;  Gaps    1;

Qy        242 TGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTA 301
              ||  ||    || ||  | | |  ||  |    | |||   | | | || | ||   |
Db       3933 TGGGTGGGTGGGTGAGTGGATGGGTGGGTAGATGATGGGTGGATGGGTGGGTGGGTGGAT 3992

Qy        302 GGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGAC 361
              |      | | ||   |||  |  ||   | | | |   |||   | | | || | |
Db       3993 GAGTGAATGGGTGGATGGATGGGTGGATAGATGGGTGAGTGGGTGGATAGATGGGTGAGT 4052

Qy        362 AGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 421
              | |  | |||  | | |  ||  | | | ||    | | | | | | | | |||
Db       4053 GAGTGAGTGGGTGGATGGGTGGGTGGATGGATAGATGTGTGGATGGGTGGGTGGATGGAT 4112

Qy        422 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 481
              | | | |   | |  ||   | | | |   | | |||   | |   | | |   |||
Db       4113 AGATGGGTGAATGGGCGGAAAGATGGGTGAATGGGTGGATGGATGAATGGATAAATGGGT 4172

Qy        482 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 541
              | | | | | | |||   | | | |   | | |||   | | | | | | | ||
Db       4173 AAATGGGTGGGTGGATGGAGAGATGGGTGAATGGGTGGGTAGGTGGATGGATGGGTGTGT 4232

Qy        542 TGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGAT 601
              | |  |||  ||  | | || ||  ||  ||   || |   | |  ||  ||
Db       4233 GGGTGAATGAATGGT--GGATGGGTGTGTGGGTGGGTGGGTGGATAGATGGGTGAATGTG 4290
```

```
Qy        602 GGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCT 661
              |  ||   | |  |  | ||||  | |||  |||  | |||| || | |  ||   | |
Db       4291 TGGATGGGTGGATAGAGGATGGGAGGATAGATGGGTGGATGGGTTGATGGATGGATGGAT 4350

Qy        662 GGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCT 721
              ||   | |     |   | |||   | |    |||    | |   | |||  |||  | |
Db       4351 GGATGGGTGTATAGATGGATGGATGGGTGAATGGGTGGGTAGATGGATAGATGGGTGGAT 4410

Qy        722 GGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCT 781
              |||  | |    ||     ||||  | | |  |||  | ||||  | | |  ||   | |
Db       4411 GGGTCGATGTATGGATGGATGGGTAGATGGGTGGGTGGATGGGTGGATGGGTGGATGGGT 4470

Qy        782 GGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTG 826
              |||  | | |  ||   | ||||  | | |   ||  |  |  ||
Db       4471 GGGTGGATGGGTGGATGGGTGGGTGGGTGGATAGGTGGATTAATG 4515


RESULT 53
US-10-027-632-143381
; Sequence 143381, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 143381
;   LENGTH: 588
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-143381

  Query Match             6.0%;  Score 60.2;  DB 13;  Length 588;
  Best Local Similarity   49.8%;  Pred. No. 6.8e-05;
  Matches  149;  Conservative    1;  Mismatches  149;  Indels    0;  Gaps    0;

Qy        235 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 294
              | |||  || ||| || | | |  |  ||   | | | ||  |||   | |   ||   |
Db        207 GATGGGTTGATTGGTGGGTGAATGGATGGATGGTTGGATCCATGGGTAGGTGAGTGGATG 266

Qy        295 GACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 354
              ||  |  ||   | | | ||   |||  |  ||   | | | | |:|   | | | ||
Db        267 GATTGGTGGATAGATGGATGGATGGATGGATGGATGGATGGATGRATGGATGGGTGG 326

Qy        355 AGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTC 414
              || ||| ||  ||   | ||| |    ||  ||   | | | |||   | |   | | |
Db        327 AGTGGATAGATGGATGGATGGATGAGCAGATGGTTGGGTTGATGGGTGGGTGAATGGATG 386

Qy        415 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 474
              | |||   | |   | | | | ||   || | | |||||   |||   | | | | | |
Db        387 GATGGTTGGATCCATGGGTAGGTGAGTTGATGGATTGCTGAGTGGGTGGGTGGATGGATG 446

Qy        475 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 533
              | ||| | | | | | | | | |||  || |   | | | | |||  || | | | | |
Db        447 GATGGACGGATGGATGGATAGATGGATTGGTGAGTGGATGGTTGGGTTGATGGATGGTT 505
```

```
RESULT 54
US-10-017-161-1673
; Sequence 1673, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1673
;   LENGTH: 13240
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(13240)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(261)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1594)..(2145)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2900)..(3044)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (3937)..(4080)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (7221)..(7364)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (7667)..(7833)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (10614)..(10746)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (12968)..(13040)
US-10-017-161-1673

  Query Match             6.0%;  Score 60.2;  DB 12;  Length 13240;
  Best Local Similarity   52.6%;  Pred. No. 4.6e-05;
  Matches  131;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy        328 CCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGG 387
              |||||  ||  ||  |||||   |         |     | ||||   | |||     ||
Db       1783 CCTGCCCGCCTGCCCACCTGCCCACCTGTCCACCTACCTGCCTGCCCACCTGCCCACCGG 1842

Qy        388 GCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 447
              || || |    |||||  | | ||| | |  |||  | | |  ||  |||  | ||| ||
Db       1843 CCTACTTGTCCACCTGCATACCTGCCCACCTGCCCACCTACCCGCCTGCTCACTTGCCTG 1902

Qy        448 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 507
              | |||   |  | |  | ||| | |  ||  |||| |   | || || |   ||||| ||
Db       1903 CCTGCCTACCTGTCCACCTGCATACCTGCCCGCCTACCCACCTGTTCACCTACCTGCCTG 1962

Qy        508 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTG 567
              |   |  | |  ||| | ||| | |  || |||||||  |||  |   | | |   | |||
Db       1963 CCCACCTGTTCACCTACCTGCCTCCCTGCCGGCCTGCCTGCCTACCTGTCTGCTTGTCTG 2022

Qy        568 CTTGCTTGC 576
                 ||  | |
```

```
Db       2023 TCTGTATTC 2031


RESULT 55
US-09-764-877-3186
; Sequence 3186, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC005
;  CURRENT APPLICATION NUMBER: US/09/764,877
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 4031
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3186
;   LENGTH: 223
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-877-3186

  Query Match             5.9%;  Score 59.4;  DB 10;  Length 223;
  Best Local Similarity   58.0%;  Pred. No. 0.00012;
  Matches  105;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy        369 CTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 428
              ||  | |||   | |   |  |  | |||    ||  |||||| | |||    |   |||
Db         43 CTCTTAGCTCAGTTGGGTGCTTAGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTAGGTGC 102

Qy        429 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 488
              ||| | |||    ||  |||||| | |||    || ||| | ||||  |    ||.|||
Db        103 TTGGTAGCTCAGTTGGGTGCTTGGTAGCTCAGGTGTTTGGTAGCTCAGTTAGGTGTTTGG 162

Qy        489 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 548
              | ||||  | |  |||||| | ||||  | |  |||||| | ||||  | |  ||||||
Db        163 TAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGGTAGCTCAGTTGGGTGCTTGA 222

Qy        549 T 549
              |
Db        223 T 223


RESULT 56
US-10-027-632-281951/c
; Sequence 281951, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 281951
;   LENGTH: 647
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-281951

  Query Match             5.9%;  Score 59.2;  DB 13;  Length 647;
  Best Local Similarity   63.9%;  Pred. No. 0.00011;
  Matches  106;  Conservative    0;  Mismatches   58;  Indels    2;  Gaps    1;

Qy        386 GGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 445
              | |||||| ||| || ||||||| |||  ||  |||||| |||||||  ||  |   ||
Db        406 GAGCTGCTTGCTTGCTTGCTTGCCTGCCTGCCTGCCTGCCTGCTTGCTTTCTTTCTCTCT 347

Qy        446 TGCTTGCTCG--CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTG 503
              | |||  ||   ||  | | ||| ||| ||  ||  | | ||| ||| |   ||  | |
Db        346 TTCTTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCCTTCTTTCTTT 287

Qy        504 CTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              ||| ||| ||| ||  | | ||| ||| ||  ||  | | ||| ||
Db        286 CTTTCTTTCTCTCTTTCTTTCTTTCTTTCTTTCTCTCTTTCTTTCT 241


RESULT 57
US-10-027-632-143383
; Sequence 143383, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 143383
;   LENGTH: 588
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-143383

  Query Match             5.9%;  Score 59;  DB 13;  Length 588;
  Best Local Similarity   49.8%;  Pred. No. 0.00013;
  Matches  149;  Conservative    0;  Mismatches  150;  Indels    0;  Gaps    0;

Qy        235 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 294
              | |||  || ||| || | | |  |  ||   | | | ||  |||   | |   ||  |
Db        207 GATGGGTTGATTGGTGGGTGAATGGATGGATGGTTGGATCCATGGGTAGGTGAGTGGATG 266

Qy        295 GACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 354
              ||  |  ||   | | | ||   |||  |  ||   | | | | | | |  | | | ||
Db        267 GATTGGTGGATAGATGGATGGATGGATGGATGGATGGATGGATAGATGGATGGGTG 326

Qy        355 AGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTC 414
              || ||| ||  ||   | ||| |    ||  ||   | | | |||   | |   | | |
Db        327 AGTGGATAGATGGATGGATGGATGAGCAGATGGTTGGGTTGATGGGTGGGTGAATGGATG 386

Qy        415 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 474
              | |||   | |   | | | | ||   || | | |||||   |||   | | | | | |
Db        387 GATGGTTGGATCCATGGGTAGGTGAGTTGATGGATTGCTGAGTGGGTGGGTGGATGGATG 446

Qy        475 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 533
              | ||| | | | | | | | | |||  || |   | | | | |||  || | | | | |
```

```
Db        447 GATGGACGGATGGATGGATAGATGGATTGGTGAGTGGATGGWTGGGTTGATGGATGGTT 505


RESULT 58
US-10-175-523-96
; Sequence 96, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
;  APPLICANT: Brockman, Jeffrey
;  APPLICANT:  Evans, David
;  APPLICANT:  Hook, Derek
;  APPLICANT:  Klimczak, Leszek
;  APPLICANT:  Laeng, Pascal
;  APPLICANT:  Palfreyman, Michael
;  APPLICANT:  Rajan, Prithi
;  TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
;  FILE REFERENCE: 3235/1J795-US3
;  CURRENT APPLICATION NUMBER: US/10/175,523
;  CURRENT FILING DATE:  2002-06-18
;  PRIOR APPLICATION NUMBER: US 60/299,151
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: US 60/317,828
;  PRIOR FILING DATE: 2001-09-07
;  PRIOR APPLICATION NUMBER: US 60/325,150
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 60/333,047
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: US 60/349,936
;  PRIOR FILING DATE: 2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/361,834
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 197
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 96
;   LENGTH: 110079
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(110079)
;   OTHER INFORMATION: where n may be a or g or c or t/u, unknown, or other
US-10-175-523-96

  Query Match             5.9%;  Score 59;  DB 14;  Length 110079;
  Best Local Similarity   60.1%;  Pred. No. 6.4e-05;
  Matches   98;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy         390 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 449
               | ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||
Db      100490 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 100549

Qy         450 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 509
               | ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||
Db      100550 TTCTTTCTTTCTTTCTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 100609

Qy         510 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAG 552
               | ||  ||  | | ||| ||| ||  ||  | | ||| || ||
Db      100610 TCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCAG 100652


RESULT 59
US-09-728-446-978/c
; Sequence 978, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668Alel Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 978
;   LENGTH: 393
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(393)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-978

  Query Match             5.9%;  Score 58.6;  DB 9;  Length 393;
  Best Local Similarity   50.6%;  Pred. No. 0.00016;
  Matches  124;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy         385 AGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 444
               || ||| |  |   |||||       |   ||| |  | ||||      |   |||||
Db         263 AGAGCTCCCAGAGAGCCTGGGGCGGGGGCTACTGCTCCCCCTGCTGTTGGTCCCTCCTGC 204

Qy         445 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGC 504
               |  ||   |||       |||| | | |||||   ||| | |||||   ||||  |
Db         203 TGCATTTNATGCTTTTGCTANTGCTGGATGGCTGGATGGCTGGGTTGCTGNATGGCTGGA 144

Qy         505 TTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGG 564
               | ||| | | |||||    || | | ||| | | |     | ||||   |    |  ||
Db         143 TGGCTGGGTTGCTGGATGCCTGGATGGCTGGGTTGGTGNATGGCTGNATGNCTGGATGGC 84

Qy         565 CTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGG 624
                 | |||||   |    ||| ||   |  ||    |||||| |||    ||| || |||
Db          83 TGGTTTGCTGNATGNCTGGATGGCTGGGNTGGNCNATGGCTGGATGNCTGGGTTGGTGGA 24

Qy         625 CTGCT 629
               || |
Db          23 TTGTT 19


RESULT 60
US-10-027-632-143382
; Sequence 143382, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 143382
;   LENGTH: 588
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-143382

  Query Match             5.9%;  Score 58.6;  DB 13;  Length 588;
  Best Local Similarity   49.5%;  Pred. No. 0.00016;
```

```
  Matches  148;  Conservative    1;  Mismatches  150;  Indels    0;  Gaps    0;

Qy        235 GCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 294
              | |||  || ||| || | | |  | ||   | | | ||  |||  | |   ||   |
Db        207 GATGGGTTGATTGGTGGGTGAATGGATGGATGGTTGGATCCATGGGTAGGTGAGTGGATG 266

Qy        295 GACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 354
              ||  |  ||   | | | ||  |||  |  ||   | | | | | | |   | | ||
Db        267 GATTGGTGGATAGATGGATGGATGGATGGATGGATGGATGGATAGATGGATGGGTG 326

Qy        355 AGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTC 414
              || ||| ||  ||   | ||| |    ||  ||   | | | |||  | |   | | |
Db        327 AGTGGATAGATGGATGGATGGATGAGCAGATGGTTGGGTTGATGGGTGGGTGAATGGATG 386

Qy        415 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 474
              | |||   | |   | | | | ||   || | | |||:|   |||   | | | | | |
Db        387 GATGGTTGGATCCATGGGTAGGTGAGTTGATGGATTGSTGAGTGGGTGGGTGGATGGATG 446

Qy        475 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 533
              | ||| | | | | | | | | ||| || |   | | | | ||| || | | | | |
Db        447 GATGGACGGATGGATGGATAGATGGATTGGTGAGTGGATGGTTGGGTTGATGGATGGTT 505


RESULT 61
US-10-032-393-47
; Sequence 47, Application US/10032393
; Publication No. US20030027286A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Gross, Molly
;  TITLE OF INVENTION: BACTERIAL PROMOTERS AND METHODS OF USE
;  FILE REFERENCE: ELITRA.010A
;  CURRENT APPLICATION NUMBER: US/10/032,393
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 60/259,434
;  PRIOR FILING DATE: 2000-12-27
;  PRIOR APPLICATION NUMBER: 09/948,993
;  PRIOR FILING DATE: 2001-09-06
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  NUMBER OF SEQ ID NOS: 68
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 47
;   LENGTH: 12733
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Vector pEPEF14
US-10-032-393-47

  Query Match             5.8%;  Score 58.2;  DB 14;  Length 12733;
  Best Local Similarity   41.7%;  Pred. No. 0.00013;
  Matches  363;  Conservative    0;  Mismatches  508;  Indels    0;  Gaps    0;

Qy         75 GGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGACCGAGGG 134
              ||| ||||||  ||||   |||| |    || ||||   || | |||| | |   | |||
Db       4498 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 4557

Qy        135 CCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGTGT 194
                 ||| |   |||||   |||     | | |||| | |||||   ||||  || | |
Db       4558 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 4617

Qy        195 GCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGG 254
              | | |   | | |   |||   ||  |   |   |   |  ||   |   |  | |||
Db       4618 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 4677

Qy        255 GACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTG 314
              |   |  ||  |   |   |  ||   |   | | ||||   |  ||   |   | |
Db       4678 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 4737

Qy        315 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTG 374
              ||||   |  ||   |   |  |  ||   |   | | ||||   |  |||  |  |
Db       4738 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 4797

Qy        375 GCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 434
              |   |  |  |||  |   |  ||  |   |   |   |  ||  |   |  |   |
Db       4798 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 4857

Qy        435 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 494
              |  ||   |  |   |  | ||  |   |   |   |  ||  |   |  |   |
Db       4858 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 4917

Qy        495 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 554
              |  ||  |   |  |  |  ||  |   |   |  |  ||  |   |  | |||
Db       4918 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 4977

Qy        555 GACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAG 614
              |   |  |||  |   |  |   ||||  ||  |   |  |  ||  |   |  |   |
Db       4978 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 5037

Qy        615 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 674
              ||  |  |||  |   |  |||  |  |||  |   |  |||  |  |||  |   |  |
Db       5038 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 5097

Qy        675 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 734
              ||  |  |||  |   |  |||     |||  |   |  |||  |  |||  |   |  |
Db       5098 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 5157

Qy        735 GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 794
              ||     |||  |   |  |||  |  |||  |   |  |||  |  |||  |   |  |
Db       5158 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 5217

Qy        795 GGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCC 854
              ||  |  |||  |   |  |||  |   |  ||  |   |  |   |  |  ||||
Db       5218 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 5277

Qy        855 GCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGGGCTGCCGGCCAGC 914
              |   |   |      |||  |  | | |  |||  |    | |  |||  |  ||   |
Db       5278 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 5337

Qy        915 TGCGTGCGAGGGCGTCCGAGGGCCAGACGAG 945
              | | | | || || |||| | ||    |||
Db       5338 GGGGGGGGGGGTCGACCGATGCCCTTGAGAG 5368


RESULT 62
US-10-032-393-8
; Sequence 8, Application US/10032393
; Publication No. US20030027286A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Gross, Molly
;  TITLE OF INVENTION: BACTERIAL PROMOTERS AND METHODS OF USE
;  FILE REFERENCE: ELITRA.010A
;  CURRENT APPLICATION NUMBER: US/10/032,393
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 60/259,434
;  PRIOR FILING DATE: 2000-12-27
;  PRIOR APPLICATION NUMBER: 09/948,993
;  PRIOR FILING DATE: 2001-09-06
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  NUMBER OF SEQ ID NOS: 68
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 12739
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Vector pEPEF1
US-10-032-393-8

  Query Match             5.8%;  Score 58.2;  DB 14;  Length 12739;
  Best Local Similarity   41.7%;  Pred. No. 0.00013;
```

```
  Matches '363;  Conservative    0;  Mismatches  508;  Indels    0;  Gaps    0;

Qy         75 GGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGACCGAGGG 134
              ||| |||||| |||| |||| |    || ||||  || | |||| | |   | |||
Db       4504 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 4563

Qy        135 CCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGACAGGGGACGGCGTGT 194
               ||| |   |||||  |||     | | |||| | |||||   ||||  || | |
Db       4564 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 4623

Qy        195 GCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGCCTGCTTGCTGAGGG 254
              | |'|   | | |     |||   ||  |   |   |   |  ||   |   |  | |||
Db       4624 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 4683

Qy        255 GACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTG 314
              |   |  ||   |   |   |  ||   |   | | ||||   |  ||   |   |  |
Db       4684 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 4743

Qy        315 AGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTG 374
               ||||   |  ||   |   |   |  ||   |   | | ||||   |  |||  |  |
Db       4744 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 4803

Qy        375 GCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 434
              |   |   | |||  |   |  ||   |   |   |  ||   |   |   |   |
Db       4804 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 4863

Qy        435 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTC 494
              |  ||   |   |   |   |  ||   |   |   |   |  ||   |   |   |
Db       4864 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 4923

Qy        495 GCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGG 554
              |  ||   |   |   |   |  ||   |   |   |   |  ||   |   |  | |||
Db       4924 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 4983

Qy        555 GACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAG 614
              |   | |||  |   |   |   ||||  ||   |   |   |  ||   |   |   |
Db       4984 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 5043

Qy        615 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 674
              ||  |  |||  |   |  |||  |  |||  |   |  |||  |  |||  |   |  |
Db       5044 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 5103

Qy        675 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 734
              ||  |  |||  |   |  |||     |||  |   |  |||  |  |||  |   |  |
Db       5104 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 5163

Qy        735 GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 794
              ||     |||  |   |  |||  |  |||  |   |  |||  |  |||  |   |  |
Db       5164 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 5223

Qy        795 GGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTAGTGGGGCC 854
              ||  |  |||  |   |  |||  |   |  ||   |   |   |   |  ||||
Db       5224 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 5283

Qy        855 GCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGGGCTGCCGGCCAGC 914
              |   |   |      |||  | | |  |||  |   |  | |  |||  |  ||   |
Db       5284 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 5343

Qy        915 TGCGTGCGAGGGCGTCCGAGGGCCAGACGAG 945
               | | | | || || |||| | ||    |||
Db       5344 GGGGGGGGGGGTCGACCGATGCCCTTGAGAG 5374


RESULT 63
US-10-017-161-1857
; Sequence 1857, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1857
;   LENGTH: 3163
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(3163)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(2963)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (556)..(655)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (658)..(669)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (671)..697)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (704)..(707)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (709)..(736)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (739)..(743)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (745)..(747)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (750)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (752)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (754)..(756)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (761)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (763)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (818)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (843)
;   OTHER INFORMATION: a, t, c, g, unknown or other
```

; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (892)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (897)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (920)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (941)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (993)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (995)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1066)..(1067)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1082)..(1084)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1105)..(1107)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1130)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1147)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1162)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1202)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1233)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1247)..(1248)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1261)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1485)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1500)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1544)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1601)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1605)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1635)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1637)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1654)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1674)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1695)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1852)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1866)..(1965)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1968)..(1980)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1982)..(2009)

Query Match 5.8%; Score 58; DB 12; Length 3163;
Best Local Similarity 43.8%; Pred. No. 0.00017;
Matches 199; Conservative 0; Mismatches 255; Indels 0; Gaps 0;

```
Qy        57 GGTGGTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGC 116
             || ||     | |  |  | | ||||||  ||||   |||| |    || ||||  ||
Db       906 GGGGGGGGGGGGGGGNGGGGTGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGG 965

Qy       117 GCGGGGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGG 176
             | |||| | |   | |||    ||| |   |||||  |||     | | |||| | |||
Db       966 GGGGGGGGGGGGGGGGGGGGGGGGGGGNGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1025

Qy       177 GGACAGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGC 236
             ||   ||||  || | | | | |   | | |     |||   ||  |   |   |
Db      1026 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNNGGGGGGGGGGGGGNNNG 1085

Qy       237 TGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 296
             ||   |   |  | ||||   |  ||   |   |   | |||   |   || | ||||
Db      1086 GGGGGGGGGGGGGGTGGGGNNNGGGGGGGGGGGGGGGGGTTGGGNGGGGGGCGGGGGGGG 1145

Qy       297 CAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAG 356
               |  ||   |   | | ||||   |  ||   |   |   | ||   |   | | |
Db      1146 GNGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGCGGNGGG 1205

Qy       357 GGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGC 416
             |||   |  |||   | |  ||  | |||  |   |       | | |  |   |
```

```
Db       1206 GGGGGGGGGGGGGGCCGCGGGGGCCCGNGGGGGGGGGGGGGGNNGGGGGGTTCGGGNGGGG 1265

Qy        417 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 476
              |  ||  || ||  |   |    |  ||   | |   | |  |  ||   | |  |
Db       1266 GTGTTTGTGTGGTTTGTTTGTTTTTTTTTTTTGTTTTGTTTTGTTTCTTTTTGTTTTTTC 1325

Qy        477 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 510
              |   || ||| ||| |||  |   |  | |  ||
Db       1326 TCCTCTTCTTTCTTCCTCTTTTCTCCTCCTCTTT 1359


RESULT 64
US-10-175-523-80
; Sequence 80, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
;  APPLICANT: Brockman, Jeffrey
;  APPLICANT:  Evans, David
;  APPLICANT:  Hook, Derek
;  APPLICANT:  Klimczak, Leszek
;  APPLICANT:  Laeng, Pascal
;  APPLICANT:  Palfreyman, Michael
;  APPLICANT:  Rajan, Prithi
;  TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
;  FILE REFERENCE: 3235/1J795-US3
;  CURRENT APPLICATION NUMBER: US/10/175,523
;  CURRENT FILING DATE:  2002-06-18
;  PRIOR APPLICATION NUMBER: US 60/299,151
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: US 60/317,828
;  PRIOR FILING DATE: 2001-09-07
;  PRIOR APPLICATION NUMBER: US 60/325,150
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 60/333,047
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: US 60/349,936
;  PRIOR FILING DATE: 2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/361,834
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 197
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 80
;   LENGTH: 90650
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-175-523-80

  Query Match             5.8%;  Score 58;  DB 14;  Length 90650;
  Best Local Similarity   54.8%;  Pred. No. 0.00011;
  Matches  115;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy        370 TGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 429
              | ||  ||| ||        | ||  ||  | | ||| ||| ||  ||  | | ||| ||
Db       2647 TCCTTTCTTCCTTTCTTCCTTTCTTCCTTTCTTCCTTTCTTCCTTTCTTCCTT 2706

Qy        430 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 489
              | ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||
Db       2707 TTCTTCCTTCCTTCCTTTCTTCCTTCCTTTCTTCCTTTCTTCCTTTCTTCCTTTCTTCCT 2766

Qy        490 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              | |  ||  | | ||| ||| ||   |  | | ||| ||| ||  ||  | | ||| ||
Db       2767 TCTTTCCTTTCTTCCTTTCTTCCTTCTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTT 2826

Qy        550 GAGGGGACAGTAGGGCTGCTTGCTTGCTAA 579
              |   || | ||    |  || | |   ||
Db       2827 TATTAGATATTATCTTTATTTACATTTCAA 2856


RESULT 65
US-10-311-455-762/c
; Sequence 762, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Det
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 762
;   LENGTH: 6171
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-762

  Query Match             5.7%;  Score 57.2;  DB 12;  Length 6171;
  Best Local Similarity   60.1%;  Pred. No. 0.00024;
  Matches   95;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        392 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 451
              ||| ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| |||
Db       5814 CTCTCTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 5755

Qy        452 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 511
              ||  ||  ||| ||| ||| ||  ||  | | ||| ||| ||| ||  | | ||| |||
Db       5754 CTTTCTTTCCTTCTTTCTTTCTTTCTTCCTTTCTTTCTTTCTCTCTCTCTTTCTTCCTTC 5695

Qy        512 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              ||  ||  ||| | | | | |    |   | | ||| ||
Db       5694 CTTCCTTCCCTCCCTCCCTCCCTCCCTCCTTCCTTCCT 5657


RESULT 66
US-10-262-552-33
; Sequence 33, Application US/10262552
; Publication No. US20030125289A1
; GENERAL INFORMATION:
;  APPLICANT: Gelb, Bruce D.
;  APPLICANT:  Tartaglia, Marco
;  TITLE OF INVENTION: NOONAN SYNDROME GENE
;  FILE REFERENCE: 2420/1J859-US1
;  CURRENT APPLICATION NUMBER: US/10/262,552
;  CURRENT FILING DATE:  2002-10-01
;  PRIOR APPLICATION NUMBER: US 60/326,532
;  PRIOR FILING DATE: 2001-10-01
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
;   LENGTH: 300000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(300000)
;   OTHER INFORMATION: where n may be a or g or c or t/u, unknown or other
US-10-262-552-33

  Query Match             5.7%;  Score 57;  DB 14;  Length 300000;
  Best Local Similarity   59.6%;  Pred. No. 0.00016;
  Matches   96;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy        386 GGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 445
              | ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||
Db      70486 GTGCTTCTTTCTTTCTTTCTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 70545
```

```
Qy        446 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 505
              | ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||
Db      70546 TCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTTCTTTCTTTCTTTCT 70605

Qy        506 TGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTT 546
              | ||| || ||  | | ||| ||| ||  ||  | | | |
Db      70606 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCCT 70646


RESULT 67
US-09-089-818B-6/c
; Sequence 6, Application US/09089818B
; Patent No. US20020164682A1
;  GENERAL INFORMATION:
;    APPLICANT:  Follettie, Maximillian
;    APPLICANT:  DeRobertis, Edward M.
;    TITLE OF INVENTION:  Mammalian Cerberus-Like Protein &
;    TITLE OF INVENTION:  Compositions
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/089,818B
;      FILING DATE:  June 3, 1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LAZAR, STEVEN R
;      REGISTRATION NUMBER:  32,618
;      REFERENCE/DOCKET NUMBER:  GI 5290A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8260
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3595 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  117..623
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  624..2402
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  2403..2699
US-09-089-818B-6

  Query Match             5.7%;  Score 56.8;  DB 10;  Length 3595;
  Best Local Similarity   62.5%;  Pred. No. 0.00031;
  Matches  105;  Conservative    0;  Mismatches   62;  Indels    1;  Gaps    1;

Qy        373 TGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG-CTTG 431
              | || |||      | ||||  ||  |||||| ||| |||  ||  |||||| || |||
Db       3282 TAGCCTGCCTCCCTGCCTGCCTGCGTGCCTGCCTGCCTGCCTGCCTGCCTGCCTTC 3223

Qy        432 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTG 491
              ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | |  || |||
Db       3222 CTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCTTTCCTTC 3163

Qy        492 CTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGG 539
              || ||  | | ||| ||  ||| ||  | ||||  ||||| |||||
Db       3162 CTTCCTTCCTTCCTTCCTCCCTCTCTATCTTGCTTCTTTGCTTGCTGG 3115


RESULT 68
US-09-803-719-1534/c
; Sequence 1534, Application US/09803719
; Publication No. US20030044783A1
; GENERAL INFORMATION:
;  APPLICANT: Williams, Lewis T.
;  APPLICANT:  Escobedo, Jaime
;  APPLICANT:  Innis, Michael A.
;  APPLICANT:  Garcia, Pablo Dominiguez
;  APPLICANT:  Sudduth-Klinger, Julie
;  APPLICANT:  Reinhard, Christoph
;  APPLICANT:  Giese, Klaus
;  APPLICANT:  Randazzo, Filippo
;  APPLICANT:  Kennedy, Giulia C.
;  APPLICANT:  Pot, David
;  APPLICANT:  Kassam, Altaf
;  APPLICANT:  Lamson, George
;  APPLICANT:  Drmanac, Radoje
;  APPLICANT:  Crkvenjakov, Radomir
;  APPLICANT:  Dickson, Mark
;  APPLICANT:  Drmanac, Snezana
;  APPLICANT:  Labat, Ivan
;  APPLICANT:  Leshkowitz, Dena
;  APPLICANT:  Kita, David
;  APPLICANT:  Garcia, Veronica
;  APPLICANT:  Jones, Lee William
;  APPLICANT:  Stache-Crain, Birgit
;  TITLE OF INVENTION: Human Genes and Gene Products
;  FILE REFERENCE: 1624.002
;  CURRENT APPLICATION NUMBER: US/09/803,719
;  CURRENT FILING DATE:  2001-03-09
;  PRIOR APPLICATION NUMBER: 60/188,609
;  PRIOR FILING DATE: 2000-03-09
;  NUMBER OF SEQ ID NOS: 2396
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1534
;   LENGTH: 412
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(412)
;   OTHER INFORMATION: n = A,T,C or G
US-09-803-719-1534

  Query Match             5.7%;  Score 56.6;  DB 11;  Length 412;
  Best Local Similarity   48.1%;  Pred. No. 0.00045;
  Matches  161;  Conservative    0;  Mismatches  174;  Indels    0;  Gaps    0;

Qy        521 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAG 580
              | ||| ||| |  | |||     | ||| | | ||     ||| |    | | |
Db        377 CGGCTGGCTCGTCCCCTGCAGGTCCTGCCGGGCGGGGGCCTGGGGTCTCCGATGGGCTCT 318

Qy        581 GGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 640
              ||  ||  ||  |  |   | | | ||   |  |||   ||||| |    |  |||||
Db        317 CGGCTGGCTCGTCCCCTGCAGGTCCTGCCGGGCGGGGGCCTGGGGTCTCCGATGGGCTCT 258

Qy        641 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCC 700
              ||   |||  |         | ||||  |   ||   |||||  |    |  ||| ||
Db        257 CGGCTGGCTCATCCCTTGTAGGTCCTGCCGGGCTGGGGCCTGGGGTCTCCGATGGGGTCT 198

Qy        701 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 760
              |||   ||| ||         |  |||||  |  |||  ||||||  |    |  |||||
Db        197 TGGCTGGCTCGTCCCCTGCAGGTCCTGCTGGGCGGGGGCCTGGGGTCTCCGATGGGCTCT 138

Qy        761 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 820
              || | ||| ||         |  |||||  |  |||   |||||  |    |  |||||
Db        137 CGGCCGGCTCGTCCCCTGCAGGTCCTGCTGGGCGGGGGCCTGGGGTCTCCGATGGGCTCT 78
```

```
Qy        821 CTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCG 855
              | ||||||| |    ||      |  |  |||| |
Db         77 CGGCTGGCTCGTCCCCTGCAGGTCCTGCTGGGCTG 43


RESULT 69
US-09-864-761-9961/c
; Sequence 9961, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 9961
;   LENGTH: 573
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO Z82201.1
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
US-09-864-761-9961

  Query Match             5.6%;  Score 56;  DB 9;  Length 573;
  Best Local Similarity   51.6%;  Pred. No. 0.00059;
  Matches  128;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy        625 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 684
              ||||  ||| ||      | ||||  ||||||      | ||||  ||||||      |
Db        572 CTGCATGTACGGAATGATGCCTGCGTGTAGGGAATGATGCCTGCATGTAGGGAATGATGC 513

Qy        685 CTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGG 744
              ||||  ||||||      | |||||  ||||||      | |||   ||||||      |
Db        512 CTGCGTGTAGGGTATGATGCCTGCTTGTAGGGAATAATGCCTGTGTGTAGGGAATGATGC 453

Qy        745 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 804
              |||   ||| ||      | ||||  |||  |      | ||||  ||||||      |
Db        452 CTGTGTGTATGGAATGATGCCTGCATGTACAGAATAATGCCTGCACGTAGGGAATGATGA 393

Qy        805 CTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGC 864
              | ||  ||||||     |||   | ||  ||      |||| ||| |   |   |||
Db        392 CCGCGTGTAGGGTATGATGCCTACATGTATGGAATGATGCTTGTGTGTGTGGAGTACTGA 333

Qy        865 TACTAGGG 872
              | ||  ||
Db        332 TGCTGTGG 325


RESULT 70
US-09-864-761-26463/c
; Sequence 26463, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
```

```
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 26463
;   LENGTH: 573
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO Z82201.1
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
;   OTHER INFORMATION: SWISSPROT HIT: P04929, EVALUE 5.00e-02
;   OTHER INFORMATION: NT HIT: gi|1426388, EVALUE 2.30e-02
;   OTHER INFORMATION: EST_HUMAN HIT: BE407615.1, EVALUE 3.50e-02
US-09-864-761-26463

  Query Match             5.6%;  Score 56;  DB 9;  Length 573;
  Best Local Similarity   51.6%;  Pred. No. 0.00059;
  Matches  128;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy        625 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 684
              ||||  ||| ||      | ||||  ||||||      | ||||  ||||||      |
Db        572 CTGCATGTACGGAATGATGCCTGCGTGTAGGGAATGATGCCTGCATGTAGGGAATGATGC 513

Qy        685 CTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGG 744
              ||||  ||||||      | |||||  ||||||      | |||   ||||||      |
Db        512 CTGCGTGTAGGGTATGATGCCTGCTTGTAGGGAATAATGCCTGTGTGTAGGGAATGATGC 453

Qy        745 CTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG 804
              |||   ||| ||      | ||||  |||  |      | ||||  ||||||      |
Db        452 CTGTGTGTATGGAATGATGCCTGCATGTACAGAATAATGCCTGCACGTAGGGAATGATGA 393

Qy        805 CTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGC 864
              | ||  ||||||     |||   | ||  ||      |||| ||| |   |     |||
Db        392 CCGCGTGTAGGGTATGATGCCTACATGTATGGAATGATGCTTGTGTGTGTGGAGTACTGA 333

Qy        865 TACTAGGG 872
              | ||  ||
Db        332 TGCTGTGG 325


RESULT 71
US-10-017-128-3/c
; Sequence 3, Application US/10017128
; Publication No. US20030124536A1
; GENERAL INFORMATION:
;  APPLICANT: McCarthy, Jeanette
;  TITLE OF INVENTION: DIAGNOSIS AND TREATMENT FOR VASCULAR DISEASE
;  FILE REFERENCE: MMI-001
;  CURRENT APPLICATION NUMBER: US/10/017,128
;  CURRENT FILING DATE:  2001-12-14
;  PRIOR APPLICATION NUMBER: US 60/306,941
;  PRIOR FILING DATE: 2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/315,572
;  PRIOR FILING DATE: 2001-08-28
;  PRIOR APPLICATION NUMBER: US 60/327,488
;  PRIOR FILING DATE: 2001-10-05
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 159095
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-017-128-3

  Query Match             5.6%;  Score 55.8;  DB 14;  Length 159095;
  Best Local Similarity   62.6%;  Pred. No. 0.00032;
  Matches   87;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy        421 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 480
              || ||| ||| ||| ||  |   ||| ||| ||   | || | ||| ||||||  ||  |
Db     104885 CTTCTTTCTTTCTCTCTCTCCCTTCCTTCCTTCATTGCTTCCTTCCTTGCTTCCTTCC 104826

Qy        481 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 540
              | ||| ||| || |||  | | ||| ||| || |||  | | ||| |||  |  || ||
Db     104825 TTCCTTCCTTCCTTGCTTCCTTCCTTCCTTCCTTGCTTCCTTCCTTCCTTCTTTCCTTGC 104766

Qy        541 CTGCTTGCTGAGGGGACAG 559
              | ||| ||     |||||
Db     104765 TTCCTTCCTTCCAAGACAG 104747


RESULT 72
US-09-728-446-435/c
; Sequence 435, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 435
;   LENGTH: 354
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(354)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-435

  Query Match             5.6%;  Score 55.6;  DB 9;  Length 354;
  Best Local Similarity   55.4%;  Pred. No. 0.00078;
  Matches   97;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        376 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 435
              | ||||  | |  |     ||||   |   |||||   |  | ||  | | ||| | | |
Db        246 CCTGCTGTTGGTCCCTGCTGCTGCATTTGCTGCTTTTGCTATTGCTGGATGGCTGGATGG 187

Qy        436 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 495
              ||||  |||| | | ||| | ||||  | ||||| | | |||||   ||| | ||| | |
Db        186 CTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGGTTGGTGG 127

Qy        496 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTG 550
              ||||  | | ||| | | | |||  | ||   | | |   |||     |   ||
Db        126 ATGGCTGGATGGCTGGATGGNTGGTTTACTNAATGGNTNTATGGNTNNATNCNTG 72


RESULT 73
US-10-027-632-170164
; Sequence 170164, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
```

```
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 170164
;   LENGTH: 489
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-170164

  Query Match             5.6%;  Score 55.6;  DB 13;  Length 489;
  Best Local Similarity   59.1%;  Pred. No. 0.00074;
  Matches   91;  Conservative    2;  Mismatches   61;  Indels    0;  Gaps    0;

Qy        396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455
              ||   ||  |  ||| ||  ||  | | ||| |||:||  || || | ||||||| || :
Db        173 CTCTTCTCTTCTCTTTCTTTCTTTCTTTCTTTCTTKCTTTCTTGCTTTCTTGCTTTCTTK 232

Qy        456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515
              ||  | | ||| |||  ||  |    | ||| ||| ||  ||  | | ||| ||| ||
Db        233 CTTTCTTTCTTTCTTTTTCCTTCCTTTYCTTTCTTYCTTTCTTTCTTTCTTTCTTTCTTT 292

Qy        516 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              ||  | | ||| ||| ||  ||  | | ||| ||
Db        293 CTTTCTTTCTTTCTTTCTTTCTTTCTTCCTTCCT 326


RESULT 74
US-10-027-632-170165
; Sequence 170165, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 170165
;   LENGTH: 489
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-170165

  Query Match             5.6%;  Score 55.6;  DB 13;  Length 489;
  Best Local Similarity   59.1%;  Pred. No. 0.00074;
  Matches   91;  Conservative    2;  Mismatches   61;  Indels    0;  Gaps    0;

Qy        396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455
              ||   ||  |  ||| ||  ||  | | ||| |||:||  || || | ||||||| || :
Db        173 CTCTTCTCTTCTCTTTCTTTCTTTCTTTCTTTCTTKCTTTCTTGCTTTCTTGCTTTCTTK 232

Qy        456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515
              ||  | | ||| |||  ||  |    | ||| ||| ||  ||  | | ||| ||| ||
Db        233 CTTTCTTTCTTTCTTTTTCCTTCCTTTYCTTTCTTYCTTTCTTTCTTTCTTTCTTTCTTT 292

Qy        516 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              ||  | | ||| ||| ||  ||  | | ||| ||
Db        293 CTTTCTTTCTTTCTTTCTTTCTTTCTTCCTTCCT 326


RESULT 75
US-10-027-632-170163
; Sequence 170163, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 170163
;   LENGTH: 978
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-170163

  Query Match             5.6%;  Score 55.6;  DB 13;  Length 978;
  Best Local Similarity   59.1%;  Pred. No. 0.00068;
  Matches   91;  Conservative    2;  Mismatches   61;  Indels    0;  Gaps    0;

Qy        396 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 455
              ||   ||  |  ||| ||  ||  | | ||| |||:||  || || | ||||||| || :
Db        173 CTCTTCTCTTCTCTTTCTTTCTTTCTTTCTTTCTTKCTTTCTTGCTTTCTTGCTTTCTTK 232

Qy        456 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 515
              ||  | | ||| |||  ||  |    | ||| ||| ||  ||  | | ||| ||| ||
Db        233 CTTTCTTTCTTTCTTTTTCCTTCCTTTYCTTTCTTYCTTTCTTTCTTTCTTTCTTTCTTT 292

Qy        516 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              ||  | | ||| ||| ||  ||  | | ||| ||
Db        293 CTTTCTTTCTTTCTTTCTTTCTTTCTTCCTTCCT 326


RESULT 76
US-10-027-632-174148
; Sequence 174148, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
```

```
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 174148
;   LENGTH: 631
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-174148

  Query Match             5.5%;  Score 55.2;  DB 13;  Length 631;
  Best Local Similarity   48.4%;  Pred. No. 0.00088;
  Matches  183;  Conservative    1;  Mismatches  189;  Indels    5;  Gaps    1;

Qy        581 GGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 640
              |||   |  | || |   | ||    ||  || | | ||   || || |    |||  ||
Db          9 GGGGGTGCTCCCCGGGGCGCTGTGCCGACGGTGGTGGTGGGAGGGTGTTCCCGGGGGCGC 68

Qy        641 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCC 700
              || ||     || | |  |   || || |    ||| ||||| || |   || | |
Db         69 TGTGCCAACGGTGGTGGGGGAAGGGTGTTCCCGGGGGTGCTGTGCCGACGGTGGTGGGGG 128

Qy        701 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 760
              || ||||    ||| ||||| || | |  | | |      || ||||    ||| |||
Db        129 GAGGGTGCTCCCCGGGGTGCTGTGCCGATGTTGGTGGGGGGAGGGTGCTCCCTGGGGTGC 188

Qy        761 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG-----CTGCTAGTAGG 815
              || || | | || | | ||   || || |    |||      |      | |       ||
Db        189 TGTGCCGATGGTGGTGGTGGGAGGGTGTTCCCCGGGGGCGCTGTGCCGACGGTGGTGGGG 248

Qy        816 GCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTG 875
              |  |  |||| || |  |||| |   | | :||| |  |   || |||| |  ||| ||
Db        249 GAAGGGTGCTCCCTGGGGTGCTGGGCAGATGRTGGTGGGGGGAGCTTGCTCCCTGGGGTG 308

Qy        876 CTGTGCAGCTGGGAGAACAGAGTAGGGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGG 935
              ||| || | |||  |   | |  ||    ||     | |||      |   ||    ||
Db        309 CTGGGCCGATGGTGGTGGGGGGAGGGTGCTCCCCGRGGGTGCTGTGCTGATGGTGGTGGG 368

Qy        936 GCCAGACGAGGACACGGG 953
              ||  |  | |  | | ||
Db        369 GCTGGTGGTGCTCCCCGG 386


RESULT 77
US-10-027-632-174149
; Sequence 174149, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 174149
;   LENGTH: 631
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-174149

  Query Match             5.5%;  Score 55.2;  DB 13;  Length 631;
  Best Local Similarity   48.4%;  Pred. No. 0.00088;
  Matches  183;  Conservative    1;  Mismatches  189;  Indels    5;  Gaps    1;

Qy        581 GGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGC 640
              |||   |  | || |   | ||    ||  || | | ||   || || |    |||  ||
Db          9 GGGGGTGCTCCCCGGGGCGCTGTGCCGACGGTGGTGGTGGGAGGGTGTTCCCGGGGGCGC 68

Qy        641 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCC 700
              || ||     || | |  |   || || |    ||| ||||| || |   || | |
Db         69 TGTGCCAACGGTGGTGGGGGAAGGGTGTTCCCGGGGGTGCTGTGCCGACGGTGGTGGGGG 128

Qy        701 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGC 760
              || ||||    ||| ||||| || | |  | | |      || ||||    ||| |||
Db        129 GAGGGTGCTCCCCGGGGTGCTGTGCCGATGTTGGTGGGGGGAGGGTGCTCCCTGGGGTGC 188

Qy        761 TGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGG-----CTGCTAGTAGG 815
              || || | | || | | ||   || || |    |||      |      | |       ||
Db        189 TGTGCCGATGGTGGTGGTGGGAGGGTGTTCCCCGGGGGCGCTGTGCCGACGGTGGTGGGG 248

Qy        816 GCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTG 875
              |  |  |||| || |  |||| |   | | :||| |  |   || |||| |  ||| ||
Db        249 GAAGGGTGCTCCCTGGGGTGCTGGGCAGATGRTGGTGGGGGGAGCTTGCTCCCTGGGGTG 308

Qy        876 CTGTGCAGCTGGGAGAACAGAGTAGGGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGG 935
              ||| || | |||  |   | |  ||    ||     | |||      |   ||    ||
Db        309 CTGGGCCGATGGTGGTGGGGGGAGGGTGCTCCCCGRGGGTGCTGTGCTGATGGTGGTGGG 368

Qy        936 GCCAGACGAGGACACGGG 953
              ||  |  | |  | | ||
Db        369 GCTGGTGGTGCTCCCCGG 386


RESULT 78
US-10-027-632-68339/c
; Sequence 68339, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68339
;   LENGTH: 439
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-68339

  Query Match             5.5%;  Score 54.8;  DB 13;  Length 439;
  Best Local Similarity   60.1%;  Pred. No. 0.0011;
```

```
  Matches   89;  Conservative    1;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461
              | ||| ||: ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db        193 TTCTTTCTYTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTCCTTCCTTCCTTCCTTCCT 134

Qy        462 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 521
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db        133 TCCTTYCTTCCTTCCTTCCTTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 74

Qy        522 TGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              | ||| ||| |   ||  ||| ||| ||
Db         73 TTCTTTCTTTCATTCTTTCCTTCTTTCT 46


RESULT 79
US-10-027-632-68340/c
; Sequence 68340, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68340
;   LENGTH: 439
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-68340

  Query Match             5.5%;  Score 54.8;  DB 13;  Length 439;
  Best Local Similarity   60.1%;  Pred. No. 0.0011;
  Matches   89;  Conservative    1;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461
              | ||| ||: ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db        193 TTCTTTCTYTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTCCTTCCTTCCTTCCTTCCT 134

Qy        462 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 521
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db        133 TCCTTYCTTCCTTCCTTCCTTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 74

Qy        522 TGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              | ||| ||| |   ||  ||| ||| ||
Db         73 TTCTTTCTTTCATTCTTTCCTTCTTTCT 46


RESULT 80
US-10-027-632-295071/c
; Sequence 295071, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 295071
;   LENGTH: 439
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-295071

  Query Match             5.5%;  Score 54.8;  DB 13;  Length 439;
  Best Local Similarity   60.1%;  Pred. No. 0.0011;
  Matches   89;  Conservative    1;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461
              | ||| ||: ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db        193 TTCTTTCTYTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTCCTTCCTTCCTTCCTTCCT 134

Qy        462 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 521
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db        133 TCCTTYCTTCCTTCCTTCCTTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 74

Qy        522 TGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              | ||| ||| |   ||  ||| ||| ||
Db         73 TTCTTTCTTTCATTCTTTCCTTCTTTCT 46


RESULT 81
US-10-027-632-295072/c
; Sequence 295072, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 295072
;   LENGTH: 439
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-295072

  Query Match             5.5%;  Score 54.8;  DB 13;  Length 439;
```

```
  Best Local Similarity   60.1%;  Pred. No. 0.0011;
  Matches   89;  Conservative    1;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461
              | ||| ||: ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db        193 TTCTTTCTYTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTCCTTCCTTCCTTCCTTCCT 134

Qy        462 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 521
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db        133 TCCTTYCTTCCTTCCTTCCTTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 74

Qy        522 TGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              | ||| ||| |   ||  ||| ||| ||
Db         73 TTCTTTCTTTCATTCTTTCCTTCTTTCT 46


RESULT 82
US-10-027-632-37032/c
; Sequence 37032, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 37032
;   LENGTH: 549
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-37032

  Query Match             5.5%;  Score 54.8;  DB 13;  Length 549;
  Best Local Similarity   60.1%;  Pred. No. 0.0011;
  Matches   89;  Conservative    1;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461
              | ||| ||: ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db        193 TTCTTTCTYTCTTTCTTTCTTTCTTTCTTTCTTTCTTCCTTCCTTCCTTCCTTCCTTCCT 134

Qy        462 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 521
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db        133 TCCTTYCTTCCTTCCTTCCTTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 74

Qy        522 TGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              | ||| ||| |   ||  ||| ||| ||
Db         73 TTCTTTCTTTCATTCTTTCCTTCTTTCT 46


RESULT 83
US-10-027-632-37033/c
; Sequence 37033, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 37033
;   LENGTH: 549
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-37033

  Query Match             5.5%;  Score 54.8;  DB 13;  Length 549;
  Best Local Similarity   60.1%;  Pred. No. 0.0011;
  Matches   89;  Conservative    1;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        402 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 461
              | ||| ||: ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db        193 TTCTTTCTYTCTTTCTTTCTTTCTTTCTTTCTTTCTTCCTTCCTTCCTTCCTTCCTTCCT 134

Qy        462 TGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCC 521
              | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  |
Db        133 TCCTTYCTTCCTTCCTTCCTTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 74

Qy        522 TGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              | ||| ||| |   ||  ||| ||| ||
Db         73 TTCTTTCTTTCATTCTTTCCTTCTTTCT 46


RESULT 84
US-10-027-632-141461/c
; Sequence 141461, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 141461
;   LENGTH: 707
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-141461
```

```
  Query Match             5.5%;  Score 54.6;  DB 13;  Length 707;
  Best Local Similarity   58.2%;  Pred. No. 0.0012;
  Matches   96;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy        361 CAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 420
              ||  | |||||  || |||   || ||| ||    ||  | | ||| ||| ||  ||  |
Db        168 CACCATGGCTGGAGGATTGAGGGTGGGGATGGGATCTTTCTTTCTTTCTTTCTTTCTTTC 109

Qy        421 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGC 480
               | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||| ||  |
Db        108 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTCTC 49

Qy        481 CTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCT 525
                 ||  ||  ||| ||  |   ||| ||| ||| ||  |   ||
Db         48 TCTCTCTCTCTCTCTCTCTCTCTTTCTTTCTCTCTCTCTCTCT 4


RESULT 85
US-10-017-161-1971
; Sequence 1971, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1971
;   LENGTH: 15203
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(15203)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(377)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (872)..(1717)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1918)..(2709)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2795)..(3118)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (14905)..(15003)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (14274)..(14373)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (14466)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (14518)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (14529)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (14616)
;   OTHER INFORMATION: a, t, c, g, unknown or other
US-10-017-161-1971

  Query Match             5.5%;  Score 54.6;  DB 12;  Length 15203;
  Best Local Similarity   50.8%;  Pred. No. 0.0008;
  Matches  156;  Conservative    0;  Mismatches  149;  Indels    2;  Gaps    1;

Qy          1 CGCCGCCCGGCATGGCGGTCCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTG 60
              |||||| |||  |   | || |||||      || |   || ||  ||| ||   ||
Db       1347 CGCCGCTGGGCTCGAGCGGCCGCCGCCTCGTCACCATGAACCCCGGAGCCGCGCCCGGGA 1406

Qy         61 GTCTCCGCGCCGCCGGGAGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCGCGG 120
              |   |     ||  | |  || |   | ||||||| ||  || |  | ||    | ||||
Db       1407 GAGCCGCGCACGGGGAGGCGGAGCGGGAGGCCCGGCGCCGCTGGAAGCGGGAGCGGGCGG 1466

Qy        121 GGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGGAC 180
              |    |    ||| |  | |||  |||| |  |||||   || ||| |  |  ||| |
Db       1467 GCGAGGGGGAAGGAGCGAGCGAGGGCGGGAGGGCGGGAGGGCGGGAGAGAGGGAGGGAAG 1526

Qy        181 AGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCTGGC 240
              |  | |  | | | |         ||      | |   || ||  | || || ||  | |
Db       1527 GAGAAAGGAAGGGAGGGAAAGAAGAAGGGCGGGCGGGTGGGTGTGTCCCGGCCCGGGGAC 1586

Qy        241 CTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGT 300
              | || | |   | || |  |  ||| ||| ||  ||| ||||| | |  ||| ||  ||
Db       1587 CGGCCTCCGCCGCGGGC--TCTGCCCGCTCGCCTGCTCGCCTGGTGGGGGAGAGGGGAGC 1644

Qy        301 AGGCCTG 307
                || | |
Db       1645 GGGGCCG 1651


RESULT 86
US-10-017-161-1403
; Sequence 1403, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1403
;   LENGTH: 1117
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1117)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(917)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (170)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (173)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (207)..(306)
;   OTHER INFORMATION: a, t, c, g, unknown or other
```

```
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (343)..(344)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (351)..(356)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (360)..(361)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (378)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (380)..(383)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (398)..(399)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (401)..(402)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (405)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (431)..(432)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (434)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (437)..(438)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (440)..(444)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (464)..(465)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (468)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (484)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (487)..(497)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (509)..(510)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (525)..(527)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (532)..(533)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (539)..(542)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (548)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (552)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (557)..(561)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (565)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (573)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (577)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (588)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (605)..(610)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (615)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (617)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (619)..(622)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (629)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (646)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (648)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (650)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (652)..(653)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
```

```
;   LOCATION: (669)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (673)..(675)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (688)..(689)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (694)..(696)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (698)

  Query Match             5.4%;  Score 54.4;  DB 12;  Length 1117;
  Best Local Similarity   39.6%;  Pred. No. 0.0012;
  Matches  112;  Conservative    0;  Mismatches  171;  Indels    0;  Gaps    0;

Qy         45 GCCGCCCCGAAAGGTGGTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCG 104
              |  |    |   || ||       |  |  ||| ||||||  ||||   |||| |    |
Db        376 GGNGNNNNGGGGGGGGGGGGGGNNGNNGGNGGGGGGGGGGGGGGGGGGGGGGGGNNGNG 435

Qy        105 GCGGGGCCCGGCGCGGGGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGC 164
              |  |     || | |||| | |   | |   .  ||| |   |||||   |
Db        436 GNNGNNNNNGGGGGGGGGGGGGGGGGGGNNGGNGGGGGGGGGGGGGGGGNGGNNNNNNNNN 495

Qy        165 GAGGGGAGCGGGGGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGC 224
              |||| |·||||    ||||  || | |   | |   | | |     |||   ||  |
Db        496 NNGGGGGGGGGGGNNGGGGGGGGGGGGGGGNNNGGGGNNGGGGGNNNNGGGGGNGGGNGGG 555

Qy        225 CTGCCTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGC 284
                  |   | ||   |   |  | ||||   |  ||   |  |          |
Db        556 GNNNNNGGGNGGGGGGGNGGGNGGGGGGGGGGNGGGGGGGGGGGGGGGGNNNNNNGGGGN 615

Qy        285 TTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGG 327
                |    ||||   |  ||   |   |  | | |    |  ||
Db        616 GNGNNNNGGGGGGNGGGGGGGGGGGGGGGGNGNGNGNNGGGGG 658


RESULT 87
US-09-728-446-190/c
; Sequence 190, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0101-USA
;  CURRENT APPLICATION NUMBER: US/09/728,446
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,270
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 1461
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 190
;   LENGTH: 345
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(345)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-446-190

  Query Match             5.4%;  Score 53.8;  DB 9;  Length 345;
  Best Local Similarity   52.1%;  Pred. No. 0.002;
  Matches  100;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy        376 CTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 435
              | ||||  | |  |     ||||   |  |||||   |  | ||  | | ||| | | |
Db        210 CCTGCTGTTGGTCCCTGCTGCTGCATTTGCTGCTTTTGCTATTGCTGGATGGCTGGATGG 151

Qy        436 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCG 495
              ||||  |||| | | ||| | ||||  | |||||  | | |||||   ||| | ||| |
Db        150 CTGGGTTGCTGGATGGCTGGATGGCTGGGTTGCTGGATGGCTGGATGGCTGGATTGGTNN 91

Qy        496 CTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGG 555
              || |  | | ||  | |  || |  ||||   |   |   || |      | ||   |
Db         90 ATGNCTGGATGGCNGGATGNCTNGTTTGCTGNATGNNTGNATGNCTGNGNCGNTGCANAG 31

Qy        556 ACAGTAGGGCTG 567
                |  | ||||
Db         30 CTGGATGCGCTG 19


RESULT 88
US-10-027-632-31508/c
; Sequence 31508, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31508
;   LENGTH: 956
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(956)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-31508

  Query Match             5.4%;  Score 53.8;  DB 13;  Length 956;
  Best Local Similarity   49.3%;  Pred. No. 0.0017;
  Matches  133;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy         58 GTGGTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGCG 117
              | ||   | |||| | |  |  |||| | ||  |  |||| |    || ||||  ||| |
Db        524 GCGGGTGCGGCGCTGGCCNGCCGGGGCCGGGNCCGGGGGGGGGGGGGGGGGGGGCGGNG 465

Qy        118 CGGGGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGG 177
              |||| | |   | |||    ||| |   |||||    ||||   | | |||| | ||||
Db        464 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGNGGAAGGGGGGGGGGGGGGGGG 405

Qy        178 GACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGCT 237
              |  ||||  || | | | | |   | |||     |||   ||  ||| |   |   |
Db        404 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGAGGGGGGGGGGGGGGGCCGGGGGGGGGGGG 345

Qy        238 GGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGAC 297
              ||   |   | ||||   |  ||   |   |   |  ||   |   |  | ||||
Db        344 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 285
```

```
Qy        298 AGTAGGCCTGCTTGCTGAGGGGACAGTAGG 327
              |  ||   |    | | ||||   || ||
Db        284 GGGGGGGGGGGGGCCGGGGGGGGGGGTGGG 255


RESULT 89
US-09-791-254-4/c
; Sequence 4, Application US/09791254
; Publication No. US20030148935A1
; GENERAL INFORMATION:
;  APPLICANT: Peymann, John A.
;  APPLICANT:  Carulli, John
;  APPLICANT:  Browning, Jeffrey
;  TITLE OF INVENTION: No. US20030148935A1el Nuclear Factor Polypeptides and Nucleic Aci
;  TITLE OF INVENTION:  Encoding Same
;  FILE REFERENCE: 15966-652 (Cura-152) US
;  CURRENT APPLICATION NUMBER: US/09/791,254
;  CURRENT FILING DATE:  2001-09-13
;  PRIOR APPLICATION NUMBER: USSN 60/184,241
;  PRIOR FILING DATE: 2000-02-23
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 2201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-791-254-4

  Query Match             5.4%;  Score 53.8;  DB 12;  Length 2201;
  Best Local Similarity   46.1%;  Pred. No. 0.0015;
  Matches  181;  Conservative    0;  Mismatches  212;  Indels    0;  Gaps    0;

Qy        463 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 522
              ||| |||    |  ||     | |||  |   |  | |  |   |    | | |     |
Db       1795 GCTGGCTCATCCCTTGTAGGTCCTGCCGGGCTGGGGCCTGGGGTCTCCGATGGGCTCTCG 1736

Qy        523 GCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGG 582
              ||| ||| |  | |||     | ||| | | ||      ||| |    | | |      |
Db       1735 GCTGGCTCGTCCCCTGCAGGTCCTGCCGGGCGGGGGGCCTGGGGTCTCCGATGGGCTCTCG 1676

Qy        583 GACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 642
              |  ||  || |   | | | || | |  ||   ||||| |    |  |||||   |
Db       1675 GCTGGCTCGTCCCCTGCAGGTCCTGCTGGGCGGGGGCCTGGGGTCTCCGATGGGCTCTCG 1616

Qy        643 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTG 702
              |   |||  |        |  ||||  |   ||   |||||  |    |  ||| ||  |
Db       1615 GCTGGCTCATCCCTTGTAGGTCCTGCCGGGCTGGGGCCTGGGGTCTCCGATGGGGTCTCG 1556

Qy        703 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTG 762
              |   ||| ||         |  |||||  |  |||  ||||||  |    |  |||||   |
Db       1555 GCTGGCTCGTCCCCTGCAGGTCCTGCTGGGCGGGGGCCTGGGGTCTCCGATGGGCTCTCG 1496

Qy        763 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCT 822
              |   ||| ||         |  |||||  |   ||   |||||  |    |  |||||   |
Db       1495 GCTGGCTCGTCCCCTGCAGGTCCTGCTGGGCTGGGGCCTGGGGTCTCCGATGGGCTCTCG 1436

Qy        823 GCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCG 855
              |||||||  |    ||      |  |  |||| |
Db       1435 GCTGGCTCGTCCCCTGCAGGTCCTGCTGGGCTG 1403


RESULT 90
US-09-791-254-1/c
; Sequence 1, Application US/09791254
; Publication No. US20030148935A1
; GENERAL INFORMATION:
;  APPLICANT: Peymann, John A.
;  APPLICANT:  Carulli, John
;  APPLICANT:  Browning, Jeffrey
;  TITLE OF INVENTION: No. US20030148935A1el Nuclear Factor Polypeptides and Nucleic Aci
;  TITLE OF INVENTION:  Encoding Same
;  FILE REFERENCE: 15966-652 (Cura-152) US
;  CURRENT APPLICATION NUMBER: US/09/791,254
;  CURRENT FILING DATE:  2001-09-13
;  PRIOR APPLICATION NUMBER: USSN 60/184,241
;  PRIOR FILING DATE: 2000-02-23
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4065
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(3726)
US-09-791-254-1

  Query Match             5.4%;  Score 53.8;  DB 12;  Length 4065;
  Best Local Similarity   46.1%;  Pred. No. 0.0014;
  Matches  181;  Conservative    0;  Mismatches  212;  Indels    0;  Gaps    0;

Qy        463 GCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCT 522
              ||| |||    |  ||     | |||  |   |  | |  |   |    | | |     |
Db       3659 GCTGGCTCATCCCTTGTAGGTCCTGCCGGGCTGGGGCCTGGGGTCTCCGATGGGCTCTCG 3600

Qy        523 GCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGG 582
              ||| ||| |  | |||     | ||| | | ||      ||| |    | | |      |
Db       3599 GCTGGCTCGTCCCCTGCAGGTCCTGCCGGGCGGGGGCCTGGGGTCTCCGATGGGCTCTCG 3540

Qy        583 GACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTG 642
              |  ||  || |   | | | || | |  ||   ||||| |    |  |||||   |
Db       3539 GCTGGCTCGTCCCCTGCAGGTCCTGCCGGGCGGGGGCCTGGGGTCTCCGATGGGCTCTCG 3480

Qy        643 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTG 702
              |   |||  |        |  ||||  |   ||   |||||  |    |  ||||||  |
Db       3479 GCTGGCTCATCCCTTGTAGGTCCTGCCGGGCTGGGGCCTGGGGTCTCCGATGGGCTCTCG 3420

Qy        703 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTG 762
              |   ||| ||        |  ||||  |  |||  ||||||  |    |  |||||   |
Db       3419 GCTGGCTCGTCCCCTGCAGGTCCTGCCGGGCGGGGGCCTGGGGTCTCCGATGGGCTCTCG 3360

Qy        763 GGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCT 822
              |   ||| ||         |  |||||  |  |||  ||||||  |    |  |||||   |
Db       3359 GCTGGCTCGTCCCCTGCAGGTCCTGCTGGGCGGGGGCCTGGGGTCTCCGATGGGCTCTCG 3300

Qy        823 GCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCG 855
              |||||||  |    ||      |  |  |||| |
Db       3299 GCTGGCTCGTCCCCTGCAGGTCCTGCTGGGCTG 3267


RESULT 91
US-09-835-232-7
; Sequence 7, Application US/09835232
; Patent No. US20020098489A1
; GENERAL INFORMATION:
;  APPLICANT: Leder, Philip
;  APPLICANT:  Leader, Benjamin
;  TITLE OF INVENTION: FORMIN-2 NUCLEIC ACIDS AND POLYPEPTIDES
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: 00383/052002
;  CURRENT APPLICATION NUMBER: US/09/835,232
;  CURRENT FILING DATE:  2001-04-12
;  PRIOR APPLICATION NUMBER: US 60/196,811
;  PRIOR FILING DATE: 2000-04-13
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 170834
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(170834)
;   OTHER INFORMATION: n= A,T,C, or G
US-09-835-232-7
```

```
  Query Match             5.4%;  Score 53.8;  DB 9;  Length 170834;
  Best Local Similarity   58.4%;  Pred. No. 0.00088;
  Matches   94;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        389 CTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 448
              ||||| |||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| |
Db      81486 CTGCTTGCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC 81545

Qy        449 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGC 508
              || ||  ||  | | ||   || ||  ||  | | ||| ||| ||  ||  | | ||| |
Db      81546 TTTCTTTCTTTCTTTCTCTTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC 81605

Qy        509 TTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
              || |||  |  | | |||  || ||   |    | ||| ||
Db      81606 TTTCTCTTTTTCTTTCTTTGTTTCTTTGTTTGTTTCTTTCT 81646


RESULT 92
US-10-027-632-270367/c
; Sequence 270367, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 270367
;   LENGTH: 502
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-270367

  Query Match             5.2%;  Score 52.4;  DB 13;  Length 502;
  Best Local Similarity   54.0%;  Pred. No. 0.0038;
  Matches  107;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy        390 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 449
              | ||  ||  | | |||  || ||| ||  | | ||| ||| ||  ||  | | ||| ||
Db        300 TTCTTTCTTTCTTTCTTTTTTTCTCTCTTTCTTTCTTTCTTTCTTTCTYTTTCTTTCT 241

Qy        450 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 509
              | ||  ||  | | ||| ||| ||  ||  ||| ||| |||  |  ||| | | ||  |
Db        240 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCCTTCTTTCTTCTTTCCTGCCTTACTGCCA 181

Qy        510 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 569
              | ||  ||  | |  || ||  ||  ||    | |||  |        | || ||||
Db        180 TCCTGCCTTTCTTCTTTCCTCCCTTCCTCCTTTCCTTCTTTCATCCTGCCTTGGCCTGCC 121

Qy        570 TGCTTGCTAAGGGGACGG 587
                  ||||  |  ||| |
Db        120 AAAGTGCTGGGATGACTG 103


RESULT 93
US-10-027-632-270368/c
; Sequence 270368, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 270368
;   LENGTH: 502
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-270368

  Query Match             5.2%;  Score 52.4;  DB 13;  Length 502;
  Best Local Similarity   54.0%;  Pred. No. 0.0038;
  Matches  107;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy        390 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 449
              | ||  ||  | | |||  || ||| ||  | | ||| ||| ||  ||  | | ||| ||
Db        300 TTCTTTCTTTCTTTCTTTTTTTCTCYCTTTCTTTCTTTCTTTCTTTCTCTCTTTCTTTCT 241

Qy        450 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 509
              | ||  ||  | | ||| ||| ||  ||  ||| ||| |||  |  ||| | | ||  |
Db        240 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCCTTCTTTCTTCTTTCCTGCCTTACTGCCA 181

Qy        510 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 569
              | ||  ||  | |  || ||  ||  ||    | |||  |          | || ||||
Db        180 TCCTGCCTTTCTTCTTTCCTCCCTTCCTCCTTTCCTTCTTTCATCCTGCCTTGGCCTGCC 121

Qy        570 TGCTTGCTAAGGGGACGG 587
                  ||||  |  ||| |
Db        120 AAAGTGCTGGGATGACTG 103


RESULT 94
US-10-027-632-183437
; Sequence 183437, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
```

```
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 183437
;   LENGTH: 553
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-183437

  Query Match             5.2%;  Score 52.4;  DB 13;  Length 553;
  Best Local Similarity   55.5%;  Pred. No. 0.0038;
  Matches  101;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        368 GCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 427
              |||| |   ||| |  |    ||  || ||  | | ||| ||| ||  ||  | | |||
Db        151 GCTGTTTTGTTGTTGTTGTTCCTTTTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 210

Qy        428 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 487
              ||| ||  ||  | | ||| ||| ||  ||  |   ||  ||  ||| ||  |   ||.
Db        211 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 270

Qy        488 CTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTG 547
              ||  ||| ||  |   ||  |||.||  ||  | | ||| ||| ||  ||  | | |||
Db        271 CTCTCTCTCTCTCTCTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 330

Qy        548 CT 549
              ||
Db        331 CT 332


RESULT 95
US-10-017-161-1483
; Sequence 1483, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1483
;   LENGTH: 3133
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(3133)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(2933)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (1)..(118)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (124)..(129)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (139)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (146)..(148)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (151)..(152)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (155)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (172)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (179)..(180)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (182)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (185)..(188)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (192)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (194)..(195)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (210)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (212)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (214)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (219)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (222)..(225)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (229)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (231)..(232)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (238)..(239)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (242)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (245)
```

```
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (248)..(260)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (265)..(270)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (272)..(276)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (278)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (282)..(283)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (287)..(290)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (302)..(306)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (311)..(313)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (323)..(334)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (336)..(341)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (345)..(363)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (365)..(368)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (370)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (374)..(375)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (379)..(385)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (387)..(388)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (390)..(393)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (395)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (397)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (400)..(402)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (409)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base

  Query Match             5.2%;  Score 52.4;  DB 12;  Length 3133;
  Best Local Similarity   17.1%;  Pred. No. 0.003;
  Matches  155;  Conservative    0;  Mismatches  752;  Indels    0;  Gaps    0;

Qy         57 GGTGGTCTCCGCGCCGCCGGGAGGGGGGCCGGGGCCCGGGGCGCCTCGGCGGGGCCCGGC 116
              || ||     | |  |  ||| |||  |  || |    ||| |    || |  |
Db        672 GGGGGGNNNNGGGGGGGGGGGGGGGNNGGGGGNGNGGNGGGGGGGGGGGGGNNGGNNNNN 731

Qy        117 GCGGGGCGCGACCGAGGGCCCCGGGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGG 176
              | || | | |   | |||    |||  |    || |   |||     | | || | | |||
Db        732 GGGGNGGGGGGNNGNGGGGGGGGGGGGGGNNNGGNGGGGGGGGGNNNGGGGGGNGNGGGGG 791

Qy        177 GGACAGGGGACGGCGTGTGCGTGCTTGTGAGACACCGGGTACGGCTGCCTGCCTGCTCGC 236
              |    ||||  || |        |  | | |     |     ||  |  |       |
Db        792 GNGGGGGGGGGGGGGGNNNNNNGGGGGGGGGGGGGGGNNNNGGGGGGGGGGNNGNGGGGG 851

Qy        237 TGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGA 296
              |    |  |     |||   |  ||  |   |   |  ||       |  ||||
Db        852 GGNNGGGGGGGNNNNNGGGGGGGGGGGNGGGNGNGGGGGGGGGGGNNNNGGGGGGGGGG 911

Qy        297 CAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTGAG 356
              |  |    |   | | ||||   | |    |   |  |  ||       |    |
Db        912 GGGGNGNGNGGGGNGNNGGGGGGGGGGGGGNNGGGNNNGGGGGNGGGGGGNGGGGGGNGG 971

Qy        357 GGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGC 416
              |||        |     ||   |   |  |||
Db        972 GGGGGGNGNNNGNGNGGGGGGGGGGGGGGGGGNNNNNNNNNNNNNNNNNNNNNNNNNNN 1031

Qy        417 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 476

Db       1032 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1091

Qy        477 TGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGC 536

Db       1092 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1151

Qy        537 TGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGC 596

Db       1152 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1211

Qy        597 CTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 656

Db       1212 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1271

Qy        657 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGG 716

Db       1272 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1331

Qy        717 CTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGG 776

Db       1332 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1391

Qy        777 CTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGC 836

Db       1392 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1451

Qy        837 TTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGA 896
                 || ||         |  |      |||     |          | || |
```

```
Db       1452 NNNNNNNNNAGGGGNNNNNNNCACGGGGGNGCCGGGNNNNNGGNNNTTNTTAAAAAAACC 1511

Qy        897 GTAGGGCTGCCGGCCAGCTGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACC 956

Db       1512 GTAAGNNCCCCNTNNGNNAGNNNTCTNNANNCCCCGATNGCCANTNGNNANCNNCCGNNN 1571

Qy        957 CGGGCCT 963

Db       1572 NNNNNTT 1578


RESULT 96
US-10-017-161-2377
; Sequence 2377, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2377
;   LENGTH: 1297
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1297)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1097)
US-10-017-161-2377

  Query Match             5.2%;  Score 52.2;  DB 12;  Length 1297;
  Best Local Similarity   44.0%;  Pred. No. 0.0038;
  Matches  265;  Conservative    0;  Mismatches  335;  Indels    2;  Gaps    1;

Qy        270 TGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCC 329

Db        102 TGCACGTATGCGTGCATGTATGCACGTATGTATGCATGTATGTGGGCATGTATGTATACA 161

Qy        330 TGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGC 389

Db        162 TGCATGTAGGCATGTATGTATACATGCATGTAGGCATGTATGCGTGTATGTATGCCTCCA 221

Qy        390 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 449

Db        222 TGTATGCATGTGTGTATACATGTATGTATGTATGCATGTGTATGCATGTGTGTATGCATG 281

Qy        450 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 509

Db        282 TGTATGCATGTATGTATGTATGCAGGTGTGTATCCATGTGTTTATACTAGTATGTATGCA 341

Qy        510 TGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCT 569

Db        342 TGTATGCATGTATACGTGTATGTATGCATGTATGCATGTATACGTGTATGTATGCATGTA 401

Qy        570 TGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCT 629

Db        402 TGCATGTATACGTGTATGTATGCATGTATGCATGTATACGTGTATGTATGCATGTATGCA 461

Qy        630 AGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCT 689

Db        462 TGTATACGTGTATGTATGCATGTATGCATGTATACGTGTATGTATGCATGTACGTATACA 521

Qy        690 AGT--AGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTG 747

Db        522 TGTGCATGCCTGTGTACATGTATGTATACATGTGTGCATGCATGTATGTATGTGTGTATG 581

Qy        748 CTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTG 807

Db        582 CATGTGTGTATGCATGCATGTATGCATGCATGTATGTGTGCATGCATGTGTGTATACATG 641

Qy        808 CTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTAC 867

Db        642 TGTGTATACGTGTCTGTGTACATGTATGTATGTATGCACGTGTGTATGTATACATGTAAG 701

Qy        868 TA 869

Db        702 TA 703


RESULT 97
US-10-311-455-985
; Sequence 985, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Det
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 985
;   LENGTH: 7597
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-985

  Query Match             5.2%;  Score 52;  DB 12;  Length 7597;
  Best Local Similarity   43.5%;  Pred. No. 0.0033;
  Matches  282;  Conservative    0;  Mismatches  365;  Indels    1;  Gaps    1;

Qy        231 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTTGCTG 290

Db       6646 GGTGGATGGTGGGATGGATGGATGGATGGATGGATTAATATATGGATGGATGGATG 6705

Qy        291 AGGGGACAGTAGGCCTGCTTGCTGAGGGGACAGTAGGCCTGCTTGCTCGCTGGCCTGCTT 350

Db       6706 AATGGACGAATGGACGGATAAATGGGTGGATGGTGGGGTGGATGTATGGATTAATATATG 6765

Qy        351 GCTGAGGGGACAGTAGGGCTGCTGGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTT 410

Db       6766 GATTGATGGATGGATGGATGGATGGATGGATGGATAGATGGATGG-GGGGTGGATG 6824

Qy        411 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 470

Db       6825 GATGGATGGATGAATACGTGGATGGATGGACGGATGGTGGGGTGGATGGATGGATGTATG 6884

Qy        471 GCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTT 530

Db       6885 AATACGTGGATGGATGGATGGTGGGGTAGATGGATGGATGAGTAGATGGATGAATGAATA 6944

Qy        531 GCTCGCTGGCCTGCTTGCTGAGGGGACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTAC 590

Db       6945 TATGGATGGATAGATGGATGGGTAGATGCGTGGATGGATGGATGAATATATGGATGGATA 7004

Qy        591 GCCTGCCTGATGGCTTGATAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 650
```

```
Db       7005 AATATATGGATGGATGGGTAAATGGGTGGATGGTGGGGTGGATGGATGGATGGATGGATG 7064

Qy        651 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTA 710

Db       7065 GATGGATGGATGGGTAGATGGGTGTATGGATGGATGAATATATGGATGGATGAATATATG 7124

Qy        711 GTAGGGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTA 770

Db       7125 GATGGATGGGTGAATGGGTGGATGGTGAGGTGGATGGATGGATGGATAGATGGGTGGATG 7184

Qy        771 GTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTT 830

Db       7185 GATGAATGAATATATGGATGGTTGGATAGATGGATGGGTAGATGAGTGGATGGATGGATG 7244

Qy        831 GCTTGCTTGCTTGCTAGTGGGGCCGCTTGCCTGCTACTAGGGCTGCTG 878

Db       7245 AGTATATGGATGGATGGATGGGTAGATGGGTGGATGGTGGGGTGGATG 7292
```

```
RESULT 98
US-09-880-107-2402
; Sequence 2402, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
;  APPLICANT: Horne, Darci T.
;  APPLICANT:  Vockley, Joseph G.
;  APPLICANT:  Scherf, Uwe
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
;  FILE REFERENCE: 44921-5028-WO
;  CURRENT APPLICATION NUMBER: US/09/880,107
;  CURRENT FILING DATE:  2001-06-14
;  PRIOR APPLICATION NUMBER: US 60/211,379
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/237,054
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 3950
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2402
;   LENGTH: 4403
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020142981A1 M80482
US-09-880-107-2402

  Query Match             5.2%;  Score 51.8;  DB 10;  Length 4403;
  Best Local Similarity   57.9%;  Pred. No. 0.004;
  Matches   92;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy         20 CCTCCGCCGGCCTCCCCTCCCCCACGCCGCCCCGAAAGGTGGTCTCCGCGCCGCCGGGAG 79

Db        123 CCCCCGGCGCACTCGCAGCTCGGGCGCCGCGCGAGCCTGTCGCCGCTATGCCTCCGCGCG 182

Qy         80 GGGGGCCGGGGCCCCGGGGCGCCTCGGCGGGGCCCGGCGCGGGGCGCGACCGAGGGCCCCG 139

Db        183 CGCCGCCTGCGCCCGGGCCCCGGCCGCCGCCCCGGGCCGCCGCCGCCACCGACACCGCCG 242

Qy        140 GGAGAACGGGGGATCGGGAAAACGCGAGGGGAGCGGGGG 178

Db        243 CGGGCGCGGGGGGCGCGGGGGGCGCGGGGGGCGCCGGCG 281
```

```
RESULT 99
US-10-017-161-1345/c
; Sequence 1345, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1345
;   LENGTH: 1360
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1360)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1160)
US-10-017-161-1345

  Query Match             5.2%;  Score 51.6;  DB 12;  Length 1360;
  Best Local Similarity   46.4%;  Pred. No. 0.0051;
  Matches  204;  Conservative    0;  Mismatches  234;  Indels    2;  Gaps    1;

Qy        555 GACAGTAGGGCTGCTTGCTTGCTAAGGGGACGGTACGCCTGCCTGATGGCTTGATAGTAG 614

Db        885 GACAGGCAGGCAGGTAGATGGATAGGCAGATGGACAGACAGGCAGACAGACAGGCAGACA 826

Qy        615 GGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 674

Db        825 GGCAGACAGACAGGCAGGCAGACGGACAGACAGATAGGCAGGCAGATGGATAGGCAGATG 766

Qy        675 GGCTGCTGGGCTGCTAGTAGGGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 734

Db        765 GACAGACGGACAGACAGGCAGACGGACAGGCAGACAGACAGGCAGACGGGCAGACAGGCA 706

Qy        735 GGCTCCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAGGGCTGCTGGGCTGCTAGTAG 794

Db        705 GACAGGCAGGCAGACAGGCAGGCAGACGGACAGGCAGACGGACAGGCAGGCAGACAGGCA 646

Qy        795 GGCTGCTGGGCTGCTAGTAGGGCTGCCTGCTGGCTTGCTTGCTTGCTTGCTAGTGGGGCC 854

Db        645 GGCAGACAGACAGGCAGGCAGGCAGACGGACAGGCAGACGGACAGGCAGGCAGACGGACA 586

Qy        855 GCTTGCCTGCTACTAGGGCTGCTGTGCAGCTGGGAGAACAGAGTAGGGCTGCCGGCCAGC 914

Db        585 GGCAGACAGGCAGACGGACAGACAGGCAGCAGACAGA--CAGGCAGGCAGGCAGGCCGAC 528

Qy        915 TGCGTGCGAGGGCGTCCGAGGGCCAGACGAGGACACGGGACCCGGGCCTCTCCCCCGCCC 974

Db        527 AGACAGGCAGGCAGATGGATAGGCAGATGGACAGATGGACAGGCAGACAGACAGGCAGAC 468

Qy        975 GGACCGCCGGGCACCCGGCC 994

Db        467 GGACAGGCAGACAGGCAGGC 448
```

```
RESULT 100
US-10-027-632-269946/c
; Sequence 269946, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
```

```
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 269946
;   LENGTH: 582
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-269946

  Query Match             5.1%;  Score 51.2;  DB 13;  Length 582;
  Best Local Similarity   55.7%;  Pred. No. 0.007;
  Matches   98;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        374 GGCTTGCTAGTAGGGCTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 433
              | ||  | |||||    |    | ||  |||     | | |   |   || || ||| ||
Db        525 GCCTCCCAAGTAGCTGGGACTACAGGTATGCACCACTACACCTGGCTATGTTTTCTTTCT 466

Qy        434 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCT 493
                || | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||| ||
Db        465 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTCT 406

Qy        494 CGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCTTGCTCGCTGGCCTGCTTGCT 549
                ||  | | ||| ||| ||  ||  | | ||| ||| ||  ||  | | ||| ||
Db        405 TTCTTYCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTTCT 350


Search completed: August 22, 2003, 00:50:43
Job time : 275.623 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 22, 2003, 00:39:18 ; Search time 15978 Seconds
(without alignments)
13801.134 Million cell updates/sec

Title: US-09-831-000-1_COPY_60966_70038
Perfect score: 9073
Sequence: 1 atggagagttccgtcggatg.....gcggccacggatcgccttga 9073

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 300 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	23	0.3	406	28	BH789496	BH789496 SALK_0297
c	2	23	0.3	564	14	CD471497	CD471497 LeukoS5_6
c	3	23	0.3	672	14	CD465526	CD465526 LeukoN1_5
	4	22	0.2	146	13	BQ312653	BQ312653 RC3-BN042
	5	22	0.2	341	9	AW314794	AW314794 11403 MAR
	6	22	0.2	364	9	AW485011	AW485011 63407 MAR
c	7	22	0.2	439	28	AQ972238	AQ972238 RPCI-23-3
	8	22	0.2	450	10	BF429531	BF429531 1067 MARC
	9	22	0.2	482	10	BF653977	BF653977 277975 MA
	10	22	0.2	487	10	BF653675	BF653675 277585 MA
	11	22	0.2	506	12	BM087765	BM087765 500577 MA
	12	22	0.2	511	10	BF042747	BF042747 BP250020B
	13	22	0.2	515	12	BI682256	BI682256 462285 MA
	14	22	0.2	526	9	AV607191	AV607191 AV607191
c	15	22	0.2	557	28	AZ463256	AZ463256 1M0271021
c	16	22	0.2	558	13	BU827158	BU827158 UK129TB07
c	17	22	0.2	563	13	BU813282	BU813282 N007E02 P
c	18	22	0.2	604	28	AZ719936	AZ719936 RPCI-24-1
c	19	22	0.2	633	13	BU874975	BU874975 V001A04 P
	20	22	0.2	636	14	CB453890	CB453890 709787 MA
c	21	22	0.2	702	14	CB166557	CB166557 IBE603019
c	22	22	0.2	709	14	CB452229	CB452229 707080 MA
c	23	22	0.2	722	28	AQ588514	AQ588514 CITBI-E1-
	24	22	0.2	752	28	AQ622034	AQ622034 HS_2005_B
	25	22	0.2	891	13	BQ686874	BQ686874 AGENCOURT
	26	22	0.2	1102	12	BM549343	BM549343 AGENCOURT
	27	22	0.2	1201	13	BX336737	BX336737 BX336737
	28	21	0.2	294	29	CC434856	CC434856 PUHDJ70TD
	29	21	0.2	365	13	BQ987865	BQ987865 QGF13G07.
	30	21	0.2	368	9	AI122904	AI122904 ok31a11.s
	31	21	0.2	375	28	AQ435960	AQ435960 HS_2258_B
c	32	21	0.2	388	28	AZ747590	AZ747590 RPCI-24-6
c	33	21	0.2	431	9	AI253067	AI253067 qz37c09.x
	34	21	0.2	448	9	AI435100	AI435100 th95c11.x
c	35	21	0.2	460	9	AW740483	AW740483 BRI10440
c	36	21	0.2	468	28	AQ247614	AQ247614 HS_2055_B
	37	21	0.2	478	12	BI073436	BI073436 kt30h02.y
	38	21	0.2	480	9	AJ466656	AJ466656 AJ466656
c	39	21	0.2	529	28	BH398988	BH398988 AG-ND-126
	40	21	0.2	531	12	BI540646	BI540646 453992 MA
c	41	21	0.2	537	28	AQ680406	AQ680406 HS_5496_B
	42	21	0.2	551	28	AZ386696	AZ386696 1M0145L17
c	43	21	0.2	553	13	BU338432	BU338432 603406611
	44	21	0.2	570	29	BZ359326	BZ359326 1d72c04.b
	45	21	0.2	599	13	BQ992662	BQ992662 QGF26E04.
c	46	21	0.2	601	29	BZ294610	BZ294610 CG1002.f1
c	47	21	0.2	617	14	CB977199	CB977199 CAB40003_
	48	21	0.2	620	29	DR18A19S	AL745916 Danio rer
c	49	21	0.2	621	29	CC372615	CC372615 PUHGJ24TD
c	50	21	0.2	649	10	BG624554	BG624554 CvHEM-037
	51	21	0.2	653	10	BF294496	BF294496 005PbG01
	52	21	0.2	656	10	BG418750	BG418750 HVSMEk002
c	53	21	0.2	684	28	BZ002130	BZ002130 oee36b07.
c	54	21	0.2	686	28	AZ278564	AZ278564 RPCI-23-1
c	55	21	0.2	693	29	BZ964704	BZ964704 PUGGI13TD
	56	21	0.2	695	13	BQ988501	BQ988501 QGF15B13.
	57	21	0.2	705	28	BZ067843	BZ067843 1kg05f07.
	58	21	0.2	708	13	BQ993877	BQ993877 QGF5L10.y
c	59	21	0.2	721	28	BH592832	BH592832 BOHFK69TR
	60	21	0.2	724	13	BQ763978	BQ763978 EBro03_SQ
c	61	21	0.2	739	9	AL965466	AL965466 AL965466
	62	21	0.2	757	10	BG596235	BG596235 EST494913
c	63	21	0.2	770	14	CB976925	CB976925 CAB40003_
	64	21	0.2	773	9	AJ514001	AJ514001 AJ514001
c	65	21	0.2	805	29	AG160514	AG160514 Pan trog1
c	66	21	0.2	848	10	BG697867	BG697867 602661083
	67	21	0.2	850	29	BX222878	BX222878 Danio rer
c	68	21	0.2	914	10	BG624651	BG624651 CvHEM-162
c	69	21	0.2	969	10	BG243673	BG243673 602357763
	70	21	0.2	1282	12	BM564204	BM564204 AGENCOURT
	71	21	0.2	1412	29	CC190145	CC190145 CH261-37F
	72	20	0.2	113	10	BE936464	BE936464 RC1-NT003
c	73	20	0.2	162	9	AW419065	AW419065 xy46g12.x
c	74	20	0.2	163	10	BF369525	BF369525 CM3-GN009
	75	20	0.2	188	10	BE000214	BE000214 MR0-BN007
c	76	20	0.2	191	10	BF327270	BF327270 MR0-BN007
	77	20	0.2	202	28	AZ308718	AZ308718 1M0012D22

	78	20	0.2	206	10	AW999124 AW999124 MR0-BN007
	79	20	0.2	209	10	BB539373 BB539373
c	80	20	0.2	216	29	DR23F24S AL975255 Danio rer
c	81	20	0.2	220	12	BI001235 BI001235 MR2-HN003
c	82	20	0.2	227	13	BQ785339 saq76c10.
c	83	20	0.2	230	14	D79467 D79467 HUM252F11B
	84	20	0.2	233	9	AV320570 AV320570 AV320570
c	85	20	0.2	235	14	H84134 H84134 yt02d12.s1
	86	20	0.2	236	28	AZ308464 AZ308464 1M0011N22
c	87	20	0.2	277	9	AA484120 AA484120 ne77g03.s
c	88	20	0.2	278	9	AA555009 AA555009 nl07c12.s
	89	20	0.2	285	29	BZ860554 BZ860554 CH240_269
c	90	20	0.2	287	10	BE702501 BE702501 RC1-NN106
c	91	20	0.2	291	9	AV108259 AV108259 AV108259
c	92	20	0.2	291	28	AQ170605 AQ170605 HS_2165_A
c	93	20	0.2	313	9	AA335616 AA335616 EST40039
	94	20	0.2	317	10	BB275640 BB275640 BB275640
c	95	20	0.2	320	10	BB501942 BB501942 BB501942
c	96	20	0.2	347	9	AA665355 AA665355 nt89e11.s
	97	20	0.2	366	10	BF477423 BF477423 nac60e12.
	98	20	0.2	371	9	AI024142 AI024142 ov73a07.s
c	99	20	0.2	373	9	AA973137 AA973137 on93b01.s
	100	20	0.2	374	9	AA664116 AA664116 ac06b06.s
c	101	20	0.2	375	12	BM376939 BM376939 EBem05_SQ
	102	20	0.2	375	28	AZ426003 AZ426003 1M0206A21
	103	20	0.2	376	9	AA782423 AA782423 ai35h06.s
	104	20	0.2	376	9	AI423650 AI423650 tf61f06.x
	105	20	0.2	378	9	AA935850 AA935850 on16b12.s
	106	20	0.2	378	9	AI624867 AI624867 ts46e08.x
	107	20	0.2	383	9	AA760870 AA760870 nz18a02.s
c	108	20	0.2	384	28	B89210 B89210 CIT-HSP-216
	109	20	0.2	387	9	AW194203 AW194203 xm07g01.x
	110	20	0.2	390	9	AI818078 AI818078 wk14d03.x
	111	20	0.2	391	12	BG991002 BG991002 RC4-HT108
	112	20	0.2	392	9	AA588766 AA588766 nn18e09.s
	113	20	0.2	393	9	AA977285 AA977285 oq52c07.s
	114	20	0.2	394	13	BY585641 BY585641
	115	20	0.2	395	14	CA951986 CA951986 iq16h05.x
	116	20	0.2	396	14	N48309 N48309 yy77h12.s1
	117	20	0.2	397	9	AA115007 AA115007 zl04c09.s
c	118	20	0.2	398	9	AI825636 AI825636 wb75a09.x
	119	20	0.2	399	9	AA846621 AA846621 aj98a05.s
	120	20	0.2	399	9	AI041688 AI041688 oy33f11.x
	121	20	0.2	400	9	AA846494 AA846494 ai85f10.s
	122	20	0.2	401	9	AI218454 AI218454 qh27a06.x
c	123	20	0.2	402	9	AA745274 AA745274 nv48g06.r
	124	20	0.2	402	9	AI337929 AI337929 qt30a02.x
	125	20	0.2	402	11	AY066884 AY066884 Schmidtea
	126	20	0.2	402	14	W96177 W96177 ze42d02.s1
	127	20	0.2	403	14	N46487 N46487 yy42c02.r1
c	128	20	0.2	404	10	BF838303 BF838303 RC1-HT097
c	129	20	0.2	404	14	BY694819 BY694819 BY694819
	130	20	0.2	404	29	CC467051 CC467051 CH240_136
c	131	20	0.2	405	10	BF838311 BF838311 RC1-HT097
	132	20	0.2	405	10	BE677604 BE677604 7d88e03.x
	133	20	0.2	406	9	AI216197 AI216197 qm41h12.x
	134	20	0.2	406	10	BE138836 BE138836 xw96c02.x
c	135	20	0.2	409	9	AI538147 AI538147 to83h04.x
	136	20	0.2	410	9	AA743889 AA743889 ob01d09.s
	137	20	0.2	411	9	AI273752 AI273752 ql47g04.x
	138	20	0.2	411	9	AI417888 AI417888 tg55a09.x
	139	20	0.2	413	9	AI569052 AI569052 tr65b10.x
	140	20	0.2	417	9	AA854228 AA854228 aj72c03.s
	141	20	0.2	417	9	AI361514 AI361514 qy36e10.x
	142	20	0.2	420	9	AI827049 AI827049 wl07f07.x
	143	20	0.2	420	28	AQ269172 AQ269172 RPCI11-76
	144	20	0.2	423	9	AI635056 AI635056 tz03g08.x
	145	20	0.2	424	9	AA700518 AA700518 ah21d06.s
c	146	20	0.2	424	9	AI210895 AI210895 m0b08a1.r
	147	20	0.2	424	9	AI302682 AI302682 qn17a03.x
c	148	20	0.2	424	12	BM742612 BM742612 K-EST0015
	149	20	0.2	426	9	AI026776 AI026776 ov97e07.x
	150	20	0.2	427	9	AA074988 AA074988 zm83a07.s
	151	20	0.2	429	9	AI076024 AI076024 oy60a09.x
	152	20	0.2	429	10	BE709518 BE709518 IL3-HT061
	153	20	0.2	430	9	AA843657 AA843657 ak08g04.s
	154	20	0.2	430	9	AI708069 AI708069 as60f02.x
	155	20	0.2	430	9	AA531550 AA531550 nj64c10.s
	156	20	0.2	431	9	AI022934 AI022934 ow65h03.s
	157	20	0.2	431	9	AI299988 AI299988 qo24d01.x
	158	20	0.2	431	9	AI476320 AI476320 ta15d03.x
	159	20	0.2	431	9	AW516670 AW516670 xq02e06.x
	160	20	0.2	432	9	AA857121 AA857121 of62d08.s
c	161	20	0.2	433	14	CB116718 CB116718 K-EST0161
	162	20	0.2	434	9	AI275033 AI275033 ql72b03.x
	163	20	0.2	435	9	AI564408 AI564408 tq74g05.x
	164	20	0.2	437	9	AI362976 AI362976 qy87h10.x
	165	20	0.2	437	14	N58925 N58925 yy62a12.s1
c	166	20	0.2	438	9	AW469126 AW469126 hc77f03.x
	167	20	0.2	439	9	AI184486 AI184486 qe26f10.x
	168	20	0.2	440	9	AA074603 AA074603 ze88d09.s
	169	20	0.2	441	9	AI085266 AI085266 oy71d02.x
	170	20	0.2	441	9	AI833224 AI833224 at76b09.x
	171	20	0.2	441	9	AI880756 AI880756 at07b07.x
	172	20	0.2	441	9	AA565705 AA565705 nk47h05.s
	173	20	0.2	443	10	BE051357 BE051357 za81f06.g
c	174	20	0.2	444	9	AV387407 AV387407 AV387407
	175	20	0.2	446	9	AI932626 AI932626 wo05d08.x
c	176	20	0.2	446	13	BU951138 BU951138 io76c08.y
	177	20	0.2	447	13	BQ345699 BQ345699 PM3-NT031
c	178	20	0.2	448	12	BM754806 BM754806 K-EST0032
	179	20	0.2	449	9	AI367133 AI367133 qk66h07.x
	180	20	0.2	450	9	AI215064 AI215064 qg52c10.x
	181	20	0.2	451	9	AI707697 AI707697 as32b12.x
	182	20	0.2	451	9	AI808882 AI808882 wf66a03.x
	183	20	0.2	452	9	AA069584 AA069584 zf75d06.s
	184	20	0.2	452	9	AA127050 AA127050 zn28e09.s
	185	20	0.2	455	9	AI142047 AI142047 oo21h12.x
	186	20	0.2	455	9	AI928253 AI928253 wo67a05.x
c	187	20	0.2	457	12	BM704059 BM704059 UI-E-CK1-
	188	20	0.2	458	9	AI304938 AI304938 qo20b03.x
	189	20	0.2	458	9	AA555175 AA555175 nl16f04.s
	190	20	0.2	459	9	AA706215 AA706215 ag93h06.s
	191	20	0.2	459	14	N39098 N39098 yy45f05.s1
	192	20	0.2	460	9	AI361323 AI361323 qy35f07.x
	193	20	0.2	460	9	AW628418 AW628418 hh94d12.x
	194	20	0.2	461	9	AW188674 AW188674 xk02a10.x
	195	20	0.2	461	9	AA400915 AA400915 zt78a04.s
	196	20	0.2	462	9	AW162485 AW162485 au67a02.x
	197	20	0.2	464	14	CA842306 CA842306 iq40b07.x
c	198	20	0.2	465	9	AA758974 AA758974 ai29b03.s
	199	20	0.2	465	9	AI022206 AI022206 ow92h01.x
c	200	20	0.2	465	10	BG621498 BG621498 602617870
	201	20	0.2	465	14	CA866874 CA866874 ir73h01.x
c	202	20	0.2	466	13	BQ824710 BQ824710 1030120F0
c	203	20	0.2	467	9	AI434092 AI434092 ti31a03.x
c	204	20	0.2	468	14	N95505 N95505 yy62a12.r1
	205	20	0.2	470	9	AI811920 AI811920 tw76c11.x
c	206	20	0.2	473	28	AZ325816 AZ325816 1M0048G10
	207	20	0.2	474	13	BX458531 BX458531 BX458531
c	208	20	0.2	475	28	AZ813493 AZ813493 2M0080K16
c	209	20	0.2	476	9	AW130519 AW130519 xf52d09.x
	210	20	0.2	476	9	AA614615 AA614615 np54a05.s
	211	20	0.2	477	9	AW440555 AW440555 hb81e11.x
c	212	20	0.2	477	10	BE121687 BE121687 894014C04
	213	20	0.2	479	14	N73466 N73466 za05e06.s1
c	214	20	0.2	479	28	AZ002337 AZ002337 RPCI-23-3
	215	20	0.2	480	9	AW829112 AW829112 ra35f02.y
	216	20	0.2	481	9	AI538542 AI538542 td08a12.x
c	217	20	0.2	481	10	BE122316 BE122316 894019B01
	218	20	0.2	482	9	AI370990 AI370990 ta07a01.x
	219	20	0.2	486	9	AA011125 AA011125 ze22e01.s
	220	20	0.2	486	9	AW070562 AW070562 xa28e07.x
	221	20	0.2	488	9	AW073795 AW073795 xb03b11.x
c	222	20	0.2	490	9	AW080432 AW080432 xe53f07.x
	223	20	0.2	492	9	AI819647 AI819647 wj87h09.x

	224	20	0.2	494	9	AI028449 AI028449 ow08b11.x
	225	20	0.2	495	9	AI028452 AI028452 ow08c02.x
	226	20	0.2	495	9	AI028454 AI028454 ow08c04.x
c	227	20	0.2	496	9	AA776810 AA776810 ah37a02.s
c	228	20	0.2	496	9	AI334229 AI334229 qq09h02.x
	229	20	0.2	496	9	AW075913 AW075913 xa81e02.x
	230	20	0.2	497	9	AI566463 AI566463 tr60b08.x
	231	20	0.2	497	9	AI872042 AI872042 wm54g11.x
c	232	20	0.2	497	14	H44373 H44373 ypl8f09.r1
	233	20	0.2	499	9	AA827558 AA827558 of33b04.s
	234	20	0.2	499	14	CA847789 CA847789 iq42b07.x
	235	20	0.2	499	28	AQ178258 AQ178258 HS_2217_B
	236	20	0.2	500	9	AI912289 AI912289 tz04e07.x
	237	20	0.2	500	13	BU948216 BU948216 io52a06.x
c	238	20	0.2	500	14	CA866609 CA866609 ir73h01.y
	239	20	0.2	502	14	N33548 N33548 yy04e03.s1
c	240	20	0.2	503	12	BI863723 BI863723 kx47f05.y
c	241	20	0.2	503	28	AQ783626 AQ783626 HS_2050_B
c	242	20	0.2	504	9	AW720788 AW720788 833006D07
	243	20	0.2	507	9	AW026493 AW026493 wv14b09.x
	244	20	0.2	507	10	BF840381 BF840381 RC1-HT097
c	245	20	0.2	507	10	BE237690 BE237690 894037E05
c	246	20	0.2	507	13	BU649218 BU649218 1112077F0
	247	20	0.2	508	9	AI696851 AI696851 wc74g07.x
	248	20	0.2	508	10	BE221255 BE221255 hu22f09.x
	249	20	0.2	511	9	AW068891 AW068891 cr38b04.x
	250	20	0.2	512	9	AI801149 AI801149 to85h03.x
	251	20	0.2	512	9	AV836101 AV836101
	252	20	0.2	513	9	AI222866 AI222866 qp40b02.x
c	253	20	0.2	513	14	W19638 W19638 zb35f09.r1
	254	20	0.2	515	9	AW474807 AW474807 xy20e02.x
	255	20	0.2	516	10	BE349252 BE349252 ht50e09.x
	256	20	0.2	516	13	BQ345577 BQ345577 PM3-NT031
	257	20	0.2	519	9	AW024881 AW024881 wu92b12.x
	258	20	0.2	519	28	AQ664010 AQ664010 HS_2277_B
c	259	20	0.2	521	12	BM693830 BM693830 UI-E-DW1-
	260	20	0.2	523	9	AI688733 AI688733 wd40g12.x
c	261	20	0.2	523	12	BI527221 BI527221 1024079D0
c	262	20	0.2	523	12	BI725948 BI725948 1031082B0
c	263	20	0.2	524	10	BE024938 BE024938 894007A04
	264	20	0.2	524	10	BE936424 BE936424 RC1-NT003
	265	20	0.2	525	9	AI708156 AI708156 as38c10.x
	266	20	0.2	525	9	AI819428 AI819428 wj90b10.x
c	267	20	0.2	525	12	BJ195233 BJ195233
	268	20	0.2	525	28	BZ019564 BZ019564 oed20e03.
	269	20	0.2	526	29	BZ910764 BZ910764 CH240_108
	270	20	0.2	527	10	BF195031 BF195031 7o94a02.x
c	271	20	0.2	527	12	BJ201858 BJ201858 BJ201858
	272	20	0.2	528	9	AA890279 AA890279 aj96g02.s
	273	20	0.2	528	10	BE675142 BE675142 7f03c04.x
	274	20	0.2	529	10	BG654564 BG654564 ib41c04.x
	275	20	0.2	532	9	AI091607 AI091607 oz94c04.x
	276	20	0.2	532	13	BQ130441 BQ130441 ij86b12.x
	277	20	0.2	534	13	BQ268745 BQ268745 ik19h08.x
	278	20	0.2	535	9	AW008532 AW008532 wv68g05.x
	279	20	0.2	536	9	AW837729 AW837729 CM1-LT004
c	280	20	0.2	538	28	AZ074757 AZ074757 RPCI-23-4
	281	20	0.2	540	10	BF090513 BF090513 RC1-NT003
	282	20	0.2	542	9	AA600845 AA600845 np45f07.s
	283	20	0.2	542	10	BF837085 BF837085 RC1-HT097
	284	20	0.2	542	10	BE855634 BE855634 7g11d06.x
c	285	20	0.2	542	12	BJ160076 BJ160076 BJ160076
c	286	20	0.2	543	12	BI496031 BI496031 df122d03.
c	287	20	0.2	543	12	BJ193003 BJ193003 BJ193003
c	288	20	0.2	544	28	AZ092083 AZ092083 RPCI-23-4
	289	20	0.2	546	10	BF840379 BF840379 RC1-HT097
	290	20	0.2	548	14	CA952502 CA952502 iq12g11.x
	291	20	0.2	549	14	CA942707 CA942707 ir66a04.x
	292	20	0.2	550	29	BZ361440 BZ361440 ie11d01.b
c	293	20	0.2	550	29	BZ365197 BZ365197 ie11d01.g
	294	20	0.2	551	10	BG058560 BG058560 naf16b11.
c	295	20	0.2	552	12	BM715216 BM715216 UI-E-CL1-
	296	20	0.2	553	9	AA716506 AA716506 zg64g07.s
	297	20	0.2	557	9	AW007056 AW007056 ws49f02.x
	298	20	0.2	557	14	CB068756 CB068756 is32e09.x
c	299	20	0.2	561	9	AV662284 AV662284 AV662284
	300	20	0.2	561	9	AW157644 AW157644 au85d06.x

ALIGNMENTS

```
RESULT 1
BH789496
LOCUS       BH789496                406 bp    DNA     linear   GSS 02-APR-2002
DEFINITION  SALK_029777.56.00.x Arabidopsis thaliana TDNA insertion lines
            Arabidopsis thaliana genomic clone SALK_029777.56.00.x, genomic
            survey sequence.
ACCESSION   BH789496
VERSION     BH789496.1  GI:19882594
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 406)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
            ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
            , Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA. This sequence lies within 300 bases of the 3' end of
            At4g17660.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .406
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="SALK_029777.56.00.x"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains  one or more TDNA insertion
                     elements. The resultant fragment for each  line was
                     directly sequenced to determine the genomic sequence at
                     the  site of insertion. Details of the protocols used can
                     be found at  http://signal.salk.edu/tdna_protocols.html"
BASE COUNT      163 a     63 c     57 g    123 t
ORIGIN

  Query Match             0.3%;  Score 23;  DB 28;  Length 406;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1354 TCTTTAAATGGAAAAACAAAAAA 1376
              |||||||||||||||||||||||
Db        189 TCTTTAAATGGAAAAACAAAAAA 211


RESULT 2
CD471497/c
LOCUS       CD471497                564 bp    mRNA    linear   EST 04-JUN-2003
DEFINITION  LeukoS5_6_A02.b1_A027 Stimulated peripheral blood leukocytes S5
            Equus caballus cDNA clone LeukoS5_6_A02_A027 3', mRNA sequence.
ACCESSION   CD471497
```

```
VERSION     CD471497.1  GI:31392765
KEYWORDS    EST.
SOURCE      Equus caballus (horse)
  ORGANISM  Equus caballus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
REFERENCE   1  (bases 1 to 564)
  AUTHORS   Vandenplas,M., Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Moore
            ,J.N., Liang,C., Sun,F., Sullivan,R., Shah,M. and Pratt,L.H.
  TITLE     An EST database from equine (Equus caballus) stimulated peripheral
            blood leukocytes
  JOURNAL   Unpublished
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
            the Human Genome Center, University of Tokyo Institute of Medical
            Science; tissue and RNA were prepared in the Department of Large
            Animal Medicine, University of Georgia; sequencing done in the
            Laboratory for Genomics and Bioinformatics, University of Georgia.
            Sequence ends have been trimmed to exclude vector and regions below
            Phred quality 16. Three-prime sequences are presented as their
            reverse complement and have been trimmed to exclude polyA.
            Seq primer: Sug3-43 (GGGGGAGGTGTGGGAGGTTTT)
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .564
                     /organism="Equus caballus"
                     /mol_type="mRNA"
                     /strain="Dartmoor Pony"
                     /db_xref="taxon:9796"
                     /clone="LeukoS5_6_A02_A027"
                     /sex="male"
                     /tissue_type="blood"
                     /cell_type="leukocytes"
                     /lab_host="DH10B-T1 phage-resistant E. coli"
                     /clone_lib="Stimulated peripheral blood leukocytes S5"
                     /note="Organ: circulatory system; Vector: pME18S-FL3;
                     Site_1: XhoI; Site_2: XhoI; The library was prepared from
                     polyA+ RNA from equine peripheral blood leukocytes
                     isolated from a healthy adult horse. The leukocytes were
                     stimulated for 4 hr with 10 ng/ml E. coli O55:B5 LPS.
                     Double-stranded cDNA was cloned unidirectionally into
                     different DraIII sites of the pME18S-FL3 vector (5-prime
                     DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG
                     ). XhoI excises the cDNA insert."
BASE COUNT      160 a    142 c    120 g    142 t
ORIGIN

  Query Match             0.3%;  Score 23;  DB 14;  Length 564;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1356 TTTAAATGGAAAAACAAAAAAAT 1378
              |||||||||||||||||||||||
Db         29 TTTAAATGGAAAAACAAAAAAAT 7


RESULT 3
CD465526/c
LOCUS       CD465526                672 bp    mRNA    linear   EST 04-JUN-2003
DEFINITION  LeukoN1_5_C05.b1_A023 Unstimulated peripheral blood leukocytes N1
            Equus caballus cDNA clone LeukoN1_5_C05_A023 3', mRNA sequence.
ACCESSION   CD465526
VERSION     CD465526.1  GI:31386794
KEYWORDS    EST.
SOURCE      Equus caballus (horse)
  ORGANISM  Equus caballus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
REFERENCE   1  (bases 1 to 672)
  AUTHORS   Vandenplas,M., Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Moore
            ,J.N., Liang,C., Sun,F., Sullivan,R., Shah,M. and Pratt,L.H.
  TITLE     An EST database from equine (Equus caballus) unstimulated
            peripheral blood leukocytes
  JOURNAL   Unpublished
COMMENT     Other_ESTs: LeukoN1_5_C05.g1_A023
            Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
            the Human Genome Center, University of Tokyo Institute of Medical
            Science; tissue and RNA were prepared in the Department of Large
            Animal Medicine, University of Georgia; sequencing done in the
            Laboratory for Genomics and Bioinformatics, University of Georgia.
            Sequence ends have been trimmed to exclude vector and regions below
            Phred quality 16. Three-prime sequences are presented as their
            reverse complement and have been trimmed to exclude polyA.
            Seq primer: Sug3 (CGACCTGCAGCTCGAGCACA)
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .672
                     /organism="Equus caballus"
                     /mol_type="mRNA"
                     /strain="Belgian draft"
                     /db_xref="taxon:9796"
                     /clone="LeukoN1_5_C05_A023"
                     /sex="female"
                     /tissue_type="blood"
                     /cell_type="leukocytes"
                     /lab_host="DH10B-T1 phage-resistant E. coli"
                     /clone_lib="Unstimulated peripheral blood leukocytes N1"
                     /note="Organ: circulatory system; Vector: pME18S-FL3;
                     Site_1: XhoI; Site_2: XhoI; The library was prepared from
                     polyA+ RNA from unstimulated equine peripheral blood
                     leukocytes isolated from a healthy adult horse.
                     Double-stranded cDNA was cloned unidirectionally into
                     different DraIII sites of the pME18S-FL3 vector (5-prime
                     DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG
                     ). XhoI excises the cDNA insert."
BASE COUNT      190 a    173 c    139 g    170 t
ORIGIN

  Query Match             0.3%;  Score 23;  DB 14;  Length 672;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1356 TTTAAATGGAAAAACAAAAAAAT 1378
              |||||||||||||||||||||||
Db        135 TTTAAATGGAAAAACAAAAAAAT 113


RESULT 4
BQ312653
LOCUS       BQ312653                146 bp    mRNA    linear   EST 16-MAY-2002
DEFINITION  RC3-BN0427-281100-013-h06 BN0427 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BQ312653
VERSION     BQ312653.1  GI:20869119
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 146)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
```

```
          ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
          Simpson,A.J.
  TITLE   Shotgun sequencing of the human transcriptome with ORF expressed
          sequence tags
  JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE 20202663
   PUBMED 10737800
COMMENT   Contact: Simpson A.J.G.
          Laboratory of Cancer Genetics
          Ludwig Institute for Cancer Research
          Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
          Brazil
          Tel: +55-11-2704922
          Fax: +55-11-2707001
          Email: asimpson@ludwig.org.br
          This sequence was derived from the FAPESP/LICR Human Cancer Genome
          Project. This entry can be seen in the following URL
          (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC3&t2=RC3-BN0427-
          281100-013-h06&t3=2000-11-28&t4=1)
          Seq primer: puc 18 forward
          High quality sequence start: 82
          High quality sequence stop: 124.
FEATURES             Location/Qualifiers
     source          1. .146
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="BN0427"
                     /note="Organ: breast_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       54 a     26 c     29 g     37 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 13;  Length 146;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6646 GTCCAATAGCTTGTCTTTAAAA 6667
              ||||||||||||||||||||||
Db         26 GTCCAATAGCTTGTCTTTAAAA 47


RESULT 5
AW314794
LOCUS       AW314794                341 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  11403 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW314794
VERSION     AW314794.1  GI:6744122
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
          Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 341)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
          Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
          ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
          Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
          Keele,J.W.
  TITLE   Sequence evaluation of four pooled-tissue normalized bovine cDNA
          libraries and construction of a gene index for cattle
  JOURNAL Genome Res. 11 (4), 626-630 (2001)
  MEDLINE 21180013
   PUBMED 11282978
COMMENT   Contact: Smith TPL
          USDA, ARS, US Meat Animal Research Center
          PO Box 166, Clay Center, NE 68933-0166, USA
          Tel: 402 762 4366
          Fax: 402 762 4390
          Email: smith@email.marc.usda.gov
          Single pass sequencing. Bases called and trimmed with phred
          v0.980904.e. Vector identified by cross_match with the -minscore 20
          and -minmatch 12 options.
          PCR PRimers
          FORWARD: AGGAAACAGCTATGACCAT
          BACKWARD: GTTTTCCCAGTCACGACG
          Plate: 7  row: O  column: 15
          Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .341
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 3BOV"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendonosus muscle, and fetal
                     longissimus muscle."
BASE COUNT       73 a     104 c     89 g     75 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 9;  Length 341;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3217 CGCTGTTCTCCGAATGAGAGTC 3238
              ||||||||||||||||||||||
Db          2 CGCTGTTCTCCGAATGAGAGTC 23


RESULT 6
AW485011
LOCUS       AW485011                364 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  63407 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW485011
VERSION     AW485011.1  GI:7055117
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
          Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 364)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
          Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
          ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
          Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
          Keele,J.W.
  TITLE   Sequence evaluation of four pooled-tissue normalized bovine cDNA
          libraries and construction of a gene index for cattle
  JOURNAL Genome Res. 11 (4), 626-630 (2001)
  MEDLINE 21180013
   PUBMED 11282978
COMMENT   Contact: Smith TPL
          USDA, ARS, US Meat Animal Research Center
          PO Box 166, Clay Center, NE 68933-0166, USA
          Tel: 402 762 4366
          Fax: 402 762 4390
          Email: smith@email.marc.usda.gov
          Single pass sequencing. Bases called and trimmed with phred
          v0.980904.e. Vector identified by cross_match with the -minscore 20
          and -minmatch 12 options.
          PCR PRimers
          FORWARD: AGGAAACAGCTATGACCAT
          BACKWARD: GTTTTCCCAGTCACGACG
          Plate: 31  row: K  column: 10
          Seq primer: ATTTAGGTGACACTATAG.
```

```
FEATURES             Location/Qualifiers
     source          1. .364
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 3BOV"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendonosus muscle, and fetal
                     longissimus muscle."
BASE COUNT       68 a    125 c     91 g     80 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 9;  Length 364;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3217 CGCTGTTCTCCGAATGAGAGTC 3238
              ||||||||||||||||||||||
Db        193 CGCTGTTCTCCGAATGAGAGTC 214


RESULT 7
AQ972238/c
LOCUS       AQ972238               439 bp    DNA     linear   GSS 28-JAN-2000
DEFINITION  RPCI-23-328I3.TV RPCI-23 Mus musculus genomic clone RPCI-23-328I3,
            genomic survey sequence.
ACCESSION   AQ972238
VERSION     AQ972238.1  GI:6802607
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 439)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 328  row: I  column: 3
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .439
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-328I3"
                     /sex="Female"
                     /lab_host="DH10B"
                     /clone_lib="RPCI-23"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).  "
BASE COUNT      134 a     84 c     93 g    126 t      2 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 28;  Length 439;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4240 TGACATTTTACTCCAAACTATT 4261
              |||||||||||||||||||||
Db         70 TGACATTTTACTCCAAACTATT 49


RESULT 8
BF429531
LOCUS       BF429531               450 bp    mRNA    linear   EST 29-NOV-2000
DEFINITION  1067 MARC BSM Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF429531
VERSION     BF429531.1  GI:11441628
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Warren,W.C., Tao,N., Allison,T., Wagner,S., Mathialagan,N., Kata,S.
            , Johnson,J., Smith,T.P.L. and Womack,J.
  TITLE     A survey of genes transcribed in bovine skeletal muscle
  JOURNAL   Unpublished
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: GGAAACAGCTATGACCATG
            BACKWARD: GTTTTCCCAGTCACGAC
            Seq primer: AATTAACCCTCACTAAAGGG.
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="Skeletal muscle"
                     /lab_host="XLOLR"
                     /clone_lib="MARC BSM"
                     /note="Vector: Uni-ZAP XR; Site_1: EcoRI; Site_2: XhoI;
                     Library obtained from Stratagene, catalog #937721. Library
                     made from skeletal muscle of a two year old Holstein cow."
BASE COUNT       98 a    132 c    123 g     97 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 10;  Length 450;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3217 CGCTGTTCTCCGAATGAGAGTC 3238
              ||||||||||||||||||||||
Db         86 CGCTGTTCTCCGAATGAGAGTC 107


RESULT 9
BF653977
LOCUS       BF653977               482 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  277975 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF653977
VERSION     BF653977.1  GI:11919109
```

```
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 482)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
   PUBMED   11282978
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 70  row: E  column: 22
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .482
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 3BOV"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendonosus muscle, and fetal
                     longissimus muscle."
BASE COUNT       86 a    169 c    125 g    102 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 10;  Length 482;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3217 CGCTGTTCTCCGAATGAGAGTC 3238
              ||||||||||||||||||||||
Db        253 CGCTGTTCTCCGAATGAGAGTC 274


RESULT 10
BF653675
LOCUS       BF653675                 487 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  277585 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF653675
VERSION     BF653675.1  GI:11918807
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
   PUBMED   11282978
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 66  row: D  column: 22
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .487
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 3BOV"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendonosus muscle, and fetal
                     longissimus muscle."
BASE COUNT       83 a    171 c    124 g    109 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 10;  Length 487;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3217 CGCTGTTCTCCGAATGAGAGTC 3238
              ||||||||||||||||||||||
Db        207 CGCTGTTCTCCGAATGAGAGTC 228


RESULT 11
BM087765
LOCUS       BM087765                 506 bp    mRNA    linear   EST 19-NOV-2001
DEFINITION  500577 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BM087765
VERSION     BM087765.1  GI:16998384
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 506)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
   PUBMED   11282978
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
```

```
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 138  row: P  column: 13
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES              Location/Qualifiers
     source           1. .506
                      /organism="Bos taurus"
                      /mol_type="mRNA"
                      /db_xref="taxon:9913"
                      /tissue_type="pooled"
                      /lab_host="DH10B"
                      /clone_lib="MARC 2BOV"
                      /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                      Library made from pooled tissue from testis, thymus,
                      semitendonosus muscle, longissimus muscle, pancreas,
                      adrenal, and endometrium."
BASE COUNT       88 a    181 c    131 g    105 t      1 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 12;  Length 506;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3217 CGCTGTTCTCCGAATGAGAGTC 3238
              ||||||||||||||||||||||
Db        390 CGCTGTTCTCCGAATGAGAGTC 411


RESULT 12
BF042747
LOCUS       BF042747                511 bp    mRNA    linear   EST 10-OCT-2000
DEFINITION  BP250020B10D12 Soares normalized bovine placenta Bos taurus cDNA
            clone BP250020B10D12 5', mRNA sequence.
ACCESSION   BF042747
VERSION     BF042747.1  GI:10759802
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinas,J., Liu,L. and Larson
            ,J.H.
  TITLE     Bovine ESTs
  JOURNAL   Unpublished
COMMENT     Contact: Lewin, H. A.
            W. M. Keck Center for Comparative and Functional Genomics
            University of Illinois at Urbana-Champaign
            340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
            61801, USA
            Tel: 217 333 5998
            Fax: 217 244 5617
            Email: h-lewin@uiuc.edu
            Funding for cattle EST sequencing was provided by the USDA National
            Research Initiative, Animal Genome Resource Grant AG 99-3205-8534
            to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED
            from Washington University Genome Center. Vector Trimmi g:
            Cross_match from Washington University Genome Center PHRAP suite.
            This sequence is vector free and at least 200 bp in length.
            PCR PRimers
            FORWARD: TAATACGACTCACTATAGGG
            BACKWARD: ATTAACCCTCACTAAAG
            Insert Length: 511   Std Error: 0.00
            Plate: BP250020B10  row: D  column: 12
            Seq primer: AGCGGATAACAATTTCACACAGGA
            High quality sequence stop: 511.
FEATURES              Location/Qualifiers
     source           1. .511
                      /organism="Bos taurus"
                      /mol_type="mRNA"
                      /db_xref="taxon:9913"
                      /clone="BP250020B10D12"
                      /sex="female"
                      /lab_host="DH10B"
                      /clone_lib="Soares normalized bovine placenta"
                      /note="Organ: placenta; Vector: pT7T3Pac; Site_1: EcoRI;
                      Site_2: NotI; The cDNA library was contributed by the
                      Soares laboratory and it  was constructed and normalized
                      as described by Bonaldo, M.F.,  Lennon, G. and Soares,
                      M.B. (1996), Genome Research 6(9): 791-806. "
BASE COUNT      104 a    155 c    149 g    103 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 10;  Length 511;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3217 CGCTGTTCTCCGAATGAGAGTC 3238
              ||||||||||||||||||||||
Db         33 CGCTGTTCTCCGAATGAGAGTC 54


RESULT 13
BI682256
LOCUS       BI682256                515 bp    mRNA    linear   EST 17-SEP-2001
DEFINITION  462285 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BI682256
VERSION     BI682256.1  GI:15635188
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 515)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
   PUBMED   11282978
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 130  row: M  column: 6
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES              Location/Qualifiers
     source           1. .515
                      /organism="Bos taurus"
                      /mol_type="mRNA"
                      /db_xref="taxon:9913"
                      /tissue_type="pooled"
                      /lab_host="DH10B"
                      /clone_lib="MARC 1BOV"
                      /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                      Library made from pooled tissue from lymph node, ovary,
                      fat, hypothalamus, and pituitary."
BASE COUNT       91 a    178 c    128 g    118 t
ORIGIN
```

```
  Query Match              0.2%;  Score 22;  DB 12;  Length 515;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3217 CGCTGTTCTCCGAATGAGAGTC 3238
              ||||||||||||||||||||||
Db        204 CGCTGTTCTCCGAATGAGAGTC 225


RESULT 14
AV607191
LOCUS       AV607191                526 bp    mRNA    linear   EST 28-NOV-2001
DEFINITION  AV607191 Bos taurus kidney fetus Bos taurus cDNA clone E1KI039D03
            5', mRNA sequence.
ACCESSION   AV607191
VERSION     AV607191.1  GI:9737564
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 526)
  AUTHORS   Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
            and Sugimoto,Y.
  TITLE     Establishment of a high throughput EST sequencing system using
            poly(A) tail-removed cDNA libraries and determination of 36,000
            bovine ESTs
  JOURNAL   Nucleic Acids Res. 29 (22), E108 (2001)
  MEDLINE   21570554
   PUBMED   11713328
COMMENT     Contact: Yoshikazu Sugimoto
            Animal Genetics Division
            Shirakawa Institute of Animal Genetics
            Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
            Tel: 81-248-25-5641
            Fax: 81-248-25-5725
            Email: kazusugi@cocoa.ocn.ne.jp
            Single pass sequencing.
            This clone was obtained from a polyA-deleted cDNA library.
FEATURES             Location/Qualifiers
     source          1. .526
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="E1KI039D03"
                     /tissue_type="kidney"
                     /dev_stage="fetus"
                     /lab_host="DH10B"
                     /clone_lib="Bos taurus kidney fetus"
                     /note="Vector: pZL1; Site_1: Sal1; Site_2: Not1; Poly A
                     was deleted from a Not1 site"
BASE COUNT       84 a    189 c    149 g    104 t
ORIGIN

  Query Match              0.2%;  Score 22;  DB 9;  Length 526;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3217 CGCTGTTCTCCGAATGAGAGTC 3238
              ||||||||||||||||||||||
Db        480 CGCTGTTCTCCGAATGAGAGTC 501


RESULT 15
AZ463256/c
LOCUS       AZ463256                557 bp    DNA     linear   GSS 04-OCT-2000
DEFINITION  1M0271O21R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0271O21 R, genomic survey sequence.
ACCESSION   AZ463256
VERSION     AZ463256.1  GI:10621381
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 557)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0271  row: O  column: 21
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 557.
FEATURES             Location/Qualifiers
     source          1. .557
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0271O21"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      186 a     92 c     65 g    213 t      1 others
ORIGIN

  Query Match              0.2%;  Score 22;  DB 28;  Length 557;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4801 TATATAATATGTATAGAATTAT 4822
              ||||||||||||||||||||||
Db        342 TATATAATATGTATAGAATTAT 321


RESULT 16
BU827158/c
LOCUS       BU827158                558 bp    mRNA    linear   EST 15-OCT-2002
DEFINITION  UK129TB07 Populus apical shoot cDNA library Populus tremula x
            Populus tremuloides cDNA 5 prime, mRNA sequence.
ACCESSION   BU827158
VERSION     BU827158.1  GI:24000610
```

```
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
  TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
            from multiple libraries
  JOURNAL   Unpublished
COMMENT     Contact: BHALERAO RUPALI R.
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: rupali.bhalerao@plantphys.umu.se.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Populus tremula x Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:47664"
                     /tissue_type="apical shoot"
                     /clone_lib="Populus apical shoot cDNA library"
BASE COUNT      156 a    139 c    114 g    149 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 13;  Length 558;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2390 AAATGGAAAACATTGTGTTTTT 2411
              ||||||||||||||||||||||
Db        553 AAATGGAAAACATTGTGTTTTT 532


RESULT 17
BU813282/c
LOCUS       BU813282                563 bp    mRNA    linear   EST 15-OCT-2002
DEFINITION  N007E02 Populus bark cDNA library Populus tremula x Populus
            tremuloides cDNA 5 prime, mRNA sequence.
ACCESSION   BU813282
VERSION     BU813282.1  GI:23969178
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 563)
  AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
  TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
            from multiple libraries
  JOURNAL   Unpublished
COMMENT     Contact: BHALERAO RUPALI R.
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: rupali.bhalerao@plantphys.umu.se.
FEATURES             Location/Qualifiers
     source          1. .563
                     /organism="Populus tremula x Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:47664"
                     /tissue_type="bark"
                     /clone_lib="Populus bark cDNA library"
BASE COUNT      158 a    132 c    116 g    157 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 13;  Length 563;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2390 AAATGGAAAACATTGTGTTTTT 2411
              ||||||||||||||||||||||
Db         511 AAATGGAAAACATTGTGTTTTT 490


RESULT 18
AZ719936/c
LOCUS       AZ719936                604 bp    DNA     linear   GSS 24-JAN-2001
DEFINITION  RPCI-24-161J16.TV RPCI-24 Mus musculus genomic clone RPCI-24-161J16
            , genomic survey sequence.
ACCESSION   AZ719936
VERSION     AZ719936.1  GI:12461108
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 604)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 161  row: J  column: 16
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .604
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-161J16"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /clone_lib="RPCI-24"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      173 a    137 c    125 g    169 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 28;  Length 604;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6769 AACAAGAAAAAATTTTTTAACC 6790
              ||||||||||||||||||||||
Db         23 AACAAGAAAAAATTTTTTAACC 2


RESULT 19
BU874975/c
LOCUS       BU874975                633 bp    mRNA    linear   EST 16-OCT-2002
DEFINITION  V001A04 Populus flower cDNA library Populus balsamifera subsp.
            trichocarpa cDNA 5 prime, mRNA sequence.
```

```
ACCESSION   BU874975
VERSION     BU874975.1  GI:24066499
KEYWORDS    EST.
SOURCE      Populus balsamifera subsp. trichocarpa
  ORGANISM  Populus balsamifera subsp. trichocarpa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 633)
  AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
  TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
            from multiple libraries
  JOURNAL   Unpublished
COMMENT     Contact: BHALERAO RUPALI R.
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: rupali.bhalerao@plantphys.umu.se.
FEATURES             Location/Qualifiers
     source          1. .633
                     /organism="Populus balsamifera subsp. trichocarpa"
                     /mol_type="mRNA"
                     /sub_species="trichocarpa"
                     /db_xref="taxon:3694"
                     /clone_lib="Populus flower cDNA library"
                     /note="Organ: flower"
BASE COUNT      189 a    141 c    129 g    174 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 13;  Length 633;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2390 AAATGGAAAACATTGTGTTTTT 2411
              ||||||||||||||||||||||
Db        532 AAATGGAAAACATTGTGTTTTT 511


RESULT 20
CB453890
LOCUS       CB453890                636 bp    mRNA    linear   EST 26-MAR-2003
DEFINITION  709787 MARC 6BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   CB453890
VERSION     CB453890.1  GI:29260272
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 636)
  AUTHORS   Smith,T.P.L., Roberts,A.J., Echternkamp,S.E., Chitko-McKown,C.G.,
            Wray,J.E. and Keele,J.W.
  TITLE     A second set of bovine ESTs from pooled-tissue normalized libraries
  JOURNAL   Unpublished
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called with phred v0.020425.c and
            trimmed with the aid of the trim_alt option. Vector identified with
            cross_match v0.990329.
            Plate: FQY8063  row: B  column: 18
            Seq primer: GTAATACGACTCACTATAGGG.
FEATURES             Location/Qualifiers
     source          1. .636
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 6BOV"
                     /note="Vector: pcDNA3.1; Site_1: EcoRI; Site_2: NotI;
                     Library made with RNA pooled from multiple tissues
                     including liver, lung, hypothalamus, pituitary, and
                     placenta/endometrium."
BASE COUNT      120 a    208 c    174 g    134 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 14;  Length 636;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3217 CGCTGTTCTCCGAATGAGAGTC 3238
              ||||||||||||||||||||||
Db        135 CGCTGTTCTCCGAATGAGAGTC 156


RESULT 21
CB166557/c
LOCUS       CB166557                702 bp    mRNA    linear   EST 30-JAN-2003
DEFINITION  IBE603019954.R1 CSEQFXN35 kidney Bos taurus cDNA, mRNA sequence.
ACCESSION   CB166557
VERSION     CB166557.1  GI:28152682
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 702)
  AUTHORS   Adelson,D.L. and Gill,C.A.
  TITLE     Bovine ESTs (Adelson and Gill)
  JOURNAL   Unpublished
COMMENT     Contact: David L. Adelson
            Animal Breeding and Genetics
            Texas A&M University
            Animal Science Dept., TAMU-2471, College Station, TX 77843-2471,
            USA
            Tel: 9798452616
            Fax: 9798456970
            Email: david.adelson@tamu.edu.
FEATURES             Location/Qualifiers
     source          1. .702
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="kidney"
                     /clone_lib="CSEQFXN35 kidney"
                     /note="Organ: kidney; Vector: pBluescript SK+; Site_1:
                     NotI; Site_2: EcoRI; sequence 5' of the insert
                     (5'-NNN. . .NNNinsert)
                     GCGAATTGGAGCTCCACCGCGGTGGCGGCCGCGGCTCGAG. Sequence 3' of
                     the inserts (AAGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAG.
                     normalized Rd 1 library, sequenced 3' with M13R primer."
BASE COUNT      154 a    174 c    226 g    148 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 14;  Length 702;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3217 CGCTGTTCTCCGAATGAGAGTC 3238
              ||||||||||||||||||||||
Db        635 CGCTGTTCTCCGAATGAGAGTC 614


RESULT 22
CB452229/c
LOCUS       CB452229                709 bp    mRNA    linear   EST 26-MAR-2003
DEFINITION  707080 MARC 6BOV Bos taurus cDNA 3', mRNA sequence.
ACCESSION   CB452229
```

```
VERSION     CB452229.1  GI:29258611
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 709)
  AUTHORS   Smith,T.P.L., Roberts,A.J., Echternkamp,S.E., Chitko-McKown,C.G.,
            Wray,J.E. and Keele,J.W.
  TITLE     A second set of bovine ESTs from pooled-tissue normalized libraries
  JOURNAL   Unpublished
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called with phred v0.020425.c and
            trimmed with the aid of the trim_alt option. Vector identified with
            cross_match v0.990329.
            Plate: FQY8063  row: B  column: 18
            Seq primer: TAGAAGGCACAGTCGAGG.
FEATURES             Location/Qualifiers
     source          1. .709
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 6BOV"
                     /note="Vector: pcDNA3.1; Site_1: EcoRI; Site_2: NotI;
                     Library made with RNA pooled from multiple tissues
                     including liver, lung, hypothalamus, pituitary, and
                     placenta/endometrium."
BASE COUNT      158 a    185 c    220 g    146 t
ORIGIN

  Query Match             0.2%;  Score 22;  DB 14;  Length 709;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3217 CGCTGTTCTCCGAATGAGAGTC 3238
              ||||||||||||||||||||||
Db        630 CGCTGTTCTCCGAATGAGAGTC 609


RESULT 23
AQ588514/c
LOCUS       AQ588514                722 bp    DNA     linear   GSS 07-JUN-1999
DEFINITION  CITBI-E1-2641D11.TR CITBI-E1 Homo sapiens genomic clone 2641D11,
            genomic survey sequence.
ACCESSION   AQ588514
VERSION     AQ588514.1  GI:5015194
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 722)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .722
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2641D11"
                     /sex="male"
                     /cell_type="sperm"
                     /clone_lib="CITBI-E1"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
BASE COUNT      192 a    151 c    136 g    242 t      1 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 28;  Length 722;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2371 TAAATAAATAAAAACAAGCAAA 2392
              ||||||||||||||||||||||
Db        341 TAAATAAATAAAAACAAGCAAA 320


RESULT 24
AQ622034
LOCUS       AQ622034                752 bp    DNA     linear   GSS 16-JUN-1999
DEFINITION  HS_2005_B1_G10_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2005 Col=19 Row=N, genomic survey
            sequence.
ACCESSION   AQ622034
VERSION     AQ622034.1  GI:5084426
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 752)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 2005  row: N  column: 19
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 752.
FEATURES             Location/Qualifiers
     source          1. .752
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=2005 Col=19 Row=N"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      218 a    159 c     97 g    278 t
```

```
ORIGIN

  Query Match             0.2%;  Score 22;  DB 28;  Length 752;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2369 ACTAAATAAATAAAAACAAGCA 2390
             ||||||||||||||||||||||
Db        275 ACTAAATAAATAAAAACAAGCA 296


RESULT 25
BQ686874
LOCUS       BQ686874                891 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8344952 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6250176
            5', mRNA sequence.
ACCESSION   BQ686874
VERSION     BQ686874.1  GI:21812190
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE.  1  (bases 1 to 891)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2392  row: n  column: 01
            High quality sequence stop: 539.
FEATURES             Location/Qualifiers
     source          1. .891
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6250176"
                     /tissue_type="ductal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_110"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed  by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of  California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      169 a    239 c    243 g    236 t      4 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 13;  Length 891;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6772 AAGAAAAAATTTTTTAACCAAA 6793
             ||||||||||||||||||||||
Db        769 AAGAAAAAATTTTTTAACCAAA 790


RESULT 26
BM549343
LOCUS       BM549343                1102 bp    mRNA    linear   EST 20-FEB-2002
DEFINITION  AGENCOURT_6557825 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5441366
            5', mRNA sequence.
ACCESSION   BM549343
VERSION     BM549343.1  GI:18784674
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1102)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1915  row: i  column: 15
            High quality sequence stop: 576.
FEATURES             Location/Qualifiers
     source          1. .1102
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5441366"
                     /tissue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_98"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      275 a    274 c    313 g    238 t      2 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 12;  Length 1102;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7799 AAGGCCAAGTGTTTGGGGGCGT 7820
             ||||||||||||||||||||||
Db        990 AAGGCCAAGTGTTTGGGGGCGT 1011


RESULT 27
BX336737
LOCUS       BX336737                1201 bp    mRNA    linear   EST 02-MAY-2003
DEFINITION  BX336737 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI031YP16 5-PRIME, mRNA sequence.
ACCESSION   BX336737
VERSION     BX336737.1  GI:30335563
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 7772.f
            Contact : Feng Liang Email : fliang@lifetech.com URL :
```

```
          http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
          Faraday Avenue Genoscope sequence ID : CS0DI031DH08QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI031YP16"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      370 a    486 c      0 g    305 t     40 others
ORIGIN

  Query Match             0.2%;  Score 22;  DB 13;  Length 1201;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6843 TCATCTCCTTCACCTCCCCTTT 6864
              ||||||||||||||||||||||
Db        331 TCATCTCCTTCACCTCCCCTTT 352


RESULT 28
CC434856
LOCUS       CC434856                294 bp    DNA     linear   GSS 20-MAY-2003
DEFINITION  PUHDJ70TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa425L20,
            genomic survey sequence.
ACCESSION   CC434856
VERSION     CC434856.1  GI:30930384
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 294)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
            ,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .294
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa425L20"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
BASE COUNT      118 a     50 c     31 g     95 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 29;  Length 294;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4868 TCAAATAATATAACACAAAAT 4888
              |||||||||||||||||||||
Db         35 TCAAATAATATAACACAAAAT 55


RESULT 29
BQ987865
LOCUS       BQ987865                365 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  QGF13G07.yg.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
            QGF13G07, mRNA sequence.
ACCESSION   BQ987865
VERSION     BQ987865.1  GI:22407399
KEYWORDS    EST.
SOURCE      Lactuca sativa
  ORGANISM  Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
            Cichorieae; Lactuca.
REFERENCE   1  (bases 1 to 365)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J., Ellison
            ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
            Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QG_CA_Contig4641, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QGF13  row: G  column: 07.
FEATURES             Location/Qualifiers
     source          1. .365
                     /organism="Lactuca sativa"
                     /mol_type="mRNA"
                     /cultivar="L.serriola"
                     /db_xref="taxon:4236"
                     /clone="QGF13G07"
                     /lab_host="E.coli"
                     /clone_lib="QG_EFGHJ lettuce serriola"
                     /note="Vector: pBRcDNASfiAB; The library was constructed
                     from 10 different sources of RNA from a single genotype.
                     Separate cDNAs were generated using primers that
                     incorporated unique 5' and 3' tags to distinguish each
                     source of RNA. cDNAs were then pooled, size-fractionated,
                     directionally cloned into a custom medium-copy vector and
                     transformations made with four size classes to minimize
                     size bias. Details of each source of RNA and library
                     construction can be obtained at http://cgpdb.ucdavis.edu/
                     TAG_LIB=QG_EFGHJ lettuce serriola
                     TAG_TISSUE=flowers environmental stress
                     TAG_SEQ=CGAATGCGGG"
BASE COUNT      130 a     58 c     97 g     80 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 13;  Length 365;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4931 AGAAATGACTACACAAACACA 4951
              |||||||||||||||||||||
Db         51 AGAAATGACTACACAAACACA 71


RESULT 30
AI122904
LOCUS       AI122904                368 bp    mRNA    linear   EST 24-SEP-1998
DEFINITION  ok31a11.s1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
            IMAGE:1509404 3', mRNA sequence.
ACCESSION   AI122904
VERSION     AI122904.1  GI:3538670
```

```
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 646    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 358.
FEATURES             Location/Qualifiers
     source          1. .368
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1509404"
                     /lab_host="DH10B"
                     /clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from five normalized
                     libraries were mixed, and ss circles were made in vitro.
                     Following HAP purification, this DNA was used as tracer in
                     a subtractive hybridization reaction. The driver was
                     PCR-amplified cDNAs from pools of 5,000 clones made from
                     the same 5 libraries. The pools consisted of the following
                     libraries and cloneIDs: Soares NbHSF pool 1:
                     309384-310919, 323208-325895 Soares Nb2HP pool 1:
                     145032-147335, 147720-148103, 148872-149255, 15002 -
                     150407, 151176-152327 Soares Nb2HF8-9W pool 1:
                     758280-760583, 772104-774407 Soares NbHPA pool 1:
                     304776-306311, 320136-322823, 326280-326663 Soares NbHOT
                     pool 1:  723720-726407, 739080-740999 Subtraction by Bento
                     Soares and M. Fatima Bonaldo."
BASE COUNT       90 a     92 c     67 g    119 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 9;  Length 368;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5316 GGGCAGCATATTTATTTTTAT 5336
              |||||||||||||||||||||
Db         16 GGGCAGCATATTTATTTTTAT 36


RESULT 31
AQ435960
LOCUS       AQ435960                375 bp    DNA     linear   GSS 31-MAR-1999
DEFINITION  HS_2258_B1_A01_T7C CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2258 Col=1 Row=B, genomic survey
            sequence.
ACCESSION   AQ435960
VERSION     AQ435960.1  GI:4547299
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 375)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 2258  row: B  column: 1
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 375.
FEATURES             Location/Qualifiers
     source          1. .375
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=2258 Col=1 Row=B"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      125 a     64 c     47 g    139 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 28;  Length 375;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4961 AATTTTATTCTCAACATATGC 4981
              |||||||||||||||||||||
Db        255 AATTTTATTCTCAACATATGC 275


RESULT 32
AZ747590/c
LOCUS       AZ747590                388 bp    DNA     linear   GSS 25-JAN-2001
DEFINITION  RPCI-24-65K4.TV RPCI-24 Mus musculus genomic clone RPCI-24-65K4,
            genomic survey sequence.
ACCESSION   AZ747590
VERSION     AZ747590.1  GI:12531584
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 388)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished
COMMENT     Other_GSSs: RPCI-24-65K4.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 65  row: K  column: 4
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .388
```

```
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-65K4"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /clone_lib="RPCI-24"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      119 a     76 c     79 g    114 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 28;  Length 388;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6012 TGAATCTATAGTTACATCTTT 6032
             |||||||||||||||||||||
Db        343 TGAATCTATAGTTACATCTTT 323


RESULT 33
AI253067/c
LOCUS       AI253067                431 bp    mRNA    linear   EST 03-FEB-1999
DEFINITION  qz37c09.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2029072 3',
            mRNA sequence.
ACCESSION   AI253067
VERSION     AI253067.1  GI:3849596
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 431)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            unknown library type
            Insert Length: 632   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 430.
FEATURES             Location/Qualifiers
     source          1. .431
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2029072"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Kid11"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo.  "
BASE COUNT       86 a    113 c    132 g    100 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 9;  Length 431;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4011 AAGGATCCTGGCCAAGGATCA 4031
              |||||||||||||||||||||
Db        293 AAGGATCCTGGCCAAGGATCA 273


RESULT 34
AI435100
LOCUS       AI435100                448 bp    mRNA    linear   EST 30-MAR-1999
DEFINITION  th95c11.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
            IMAGE:2126420 3', mRNA sequence.
ACCESSION   AI435100
VERSION     AI435100.1  GI:4300814
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 448)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 672   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 440.
FEATURES             Location/Qualifiers
     source          1. .448
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2126420"
                     /lab_host="DH10B"
                     /clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from five normalized
                     libraries were mixed, and ss circles were made in vitro.
                     Following HAP purification, this DNA was used as tracer in
                     a subtractive hybridization reaction. The driver was
                     PCR-amplified cDNAs from pools of 5,000 clones made from
                     the same 5 libraries. The pools consisted of the following
                     libraries and cloneIDs: Soares NbHSF pool 1:
                     309384-310919, 323208-325895 Soares Nb2HP pool 1:
                     145032-147335, 147720-148103, 148872-149255, 15002 -
                     150407, 151176-152327 Soares Nb2HF8-9W pool 1:
                     758280-760583, 772104-774407 Soares NbHPA pool 1:
                     304776-306311, 320136-322823, 326280-326663 Soares NbHOT
                     pool 1:  723720-726407, 739080-740999 Subtraction by Bento
                     Soares and M. Fatima Bonaldo."
BASE COUNT      105 a    115 c     78 g    150 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 9;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5316 GGGCAGCATATTTATTTTTAT 5336
              |||||||||||||||||||||
Db         16 GGGCAGCATATTTATTTTTAT 36


RESULT 35
AW740483/c
LOCUS       AW740483                460 bp    mRNA    linear   EST 27-APR-2000
DEFINITION  BRI10440 Biomphalaria glabrata (BS-90)-unexposed Lambda Zap Library
            Biomphalaria glabrata cDNA clone RBGIF73TR, mRNA sequence.
ACCESSION   AW740483
VERSION     AW740483.1  GI:7651576
KEYWORDS    EST.
```

```
SOURCE      Biomphalaria glabrata (bloodfluke planorb)
  ORGANISM  Biomphalaria glabrata
            Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora
            ; Lymnaeoidea; Planorbidae; Biomphalaria.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Raghavan,N., Miller,A.N., Gardner,M., FitzGerald,P.C., Kerlavage
            ,A.R., Johnston,D.A., Lewis,F.A. and Knight,M.
  TITLE     Comparative gene analysis of Biomphalaria glabrata hemocytes pre-
            and post-exposure to miracidia of Schistosoma mansoni
  JOURNAL   Mol. Biochem. Parasitol. 126 (2), 181-191 (2003)
  MEDLINE   22503446
   PUBMED   12615317
COMMENT     Contact: Raghavan N
            Biomedical Research Institute
            12111 Parklawn Dr., Rockville, MD 20852, USA
            Tel: 301-881-3300 ext.128
            Fax: 301-770-4756
            Email: nkr@helix.nih.gov, snailsrule@aol.com.
FEATURES             Location/Qualifiers
     source          1. .460
                     /organism="Biomphalaria glabrata"
                     /mol_type="mRNA"
                     /strain="BS-90"
                     /db_xref="taxon:6526"
                     /clone="RBGIF73TR"
                     /sex="hermaphrodite"
                     /cell_type="Hemocyte"
                     /lab_host="Laboratory host"
                     /clone_lib="Biomphalaria glabrata (BS-90)-unexposed Lambda
                     Zap Library"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Total RNA was isolated from the hemocytes of
                     unexposed Biomphalaria glabrata (BS-90) snails and first
                     strand cDNA synthesized using an oligo-dT primer-linker
                     (XhoI). Second strand synthesis was followed by the
                     ligation of EcoRI adaptors. Following digestion with XhoI,
                     the  completed, directional cDNA was cloned into Uni-ZAP
                     XR Phagemid  vector by Stratgene.                    "
BASE COUNT      156 a     64 c     53 g    187 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 9;  Length 460;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2372 AAATAAATAAAAACAAGCAAA 2392
              |||||||||||||||||||||
Db        263 AAATAAATAAAAACAAGCAAA 243


RESULT 36
AQ247614/c
LOCUS       AQ247614                468 bp    DNA     linear   GSS 06-OCT-1998
DEFINITION  HS_2055_B2_E08_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2055 Col=16 Row=J, genomic survey
            sequence.
ACCESSION   AQ247614
VERSION     AQ247614.1  GI:3697796
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 468)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2055  row: J  column: 16
            Class: BAC ends
            High quality sequence stop: 468.
FEATURES             Location/Qualifiers
     source          1. .468
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=2055 Col=16 Row=J"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      126 a    105 c     86 g    149 t      2 others
ORIGIN

  Query Match             0.2%;  Score 21;  DB 28;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7488 ACCTAAATAAGCTCACAGAAG 7508
              |||||||||||||||||||||
Db        196 ACCTAAATAAGCTCACAGAAG 176


RESULT 37
BI073436
LOCUS       BI073436                478 bp    mRNA    linear   EST 19-JUN-2001
DEFINITION  kt30h02.y1 Strongyloides ratti L2 pAMP1 v1 Chiapelli McCarter
            Strongyloides ratti cDNA 5', mRNA sequence.
ACCESSION   BI073436
VERSION     BI073436.1  GI:14494056
KEYWORDS    EST.
SOURCE      Strongyloides ratti
  ORGANISM  Strongyloides ratti
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Panagrolaimoidea; Strongyloididae; Strongyloides.
REFERENCE   1  (bases 1 to 478)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.
            , Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.
            , Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagareishvili,R.
            , Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
            ,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.
            , Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
            Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished
COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            The library was constructed by Brandi Chiapelli and Dr. James
            McCarter (bchiapel@watson.wustl.edu & jmccarte@watson.wustl.edu) at
            Washington University, St. Louis. DNA Sequencing by: Washington
            University Genome Sequencing Center St. Louis.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 373.
FEATURES             Location/Qualifiers
     source          1. .478
                     /organism="Strongyloides ratti"
                     /mol_type="mRNA"
                     /db_xref="taxon:34506"
                     /dev_stage="L2"
```

```
                     /lab_host="DH10B"
                     /clone_lib="Strongyloides ratti L2 pAMP1 v1 Chiapelli
                     McCarter"
                     /note="Vector: pAMP1 (Gibco); The library was constructed
                     by Brandi Chiapelli and Dr. James  McCarter at Washington
                     University, St. Louis. The cDNA was  made by using
                     Dynabead oligo-dT priming (Dynal). PCR based  library
                     using a modified protocol from the SMART PCR cDNA
                     Synthesis Kit from Clontech. Directionally cloned into the
                     UDG sites of pAMP1. Nematodes were provided by Dr. Mark
                     Viney  of Bristol, UK."
BASE COUNT      168 a     67 c     57 g    186 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 12;  Length 478;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4662 AAGTGGTATTATTTTACAACA 4682
              |||||||||||||||||||||
Db        391 AAGTGGTATTATTTTACAACA 411


RESULT 38
AJ466656
LOCUS       AJ466656                480 bp    mRNA    linear   EST 24-MAY-2002
DEFINITION  AJ466656 S00008 Hordeum vulgare cDNA clone S0000800177G09F1, mRNA
            sequence.
ACCESSION   AJ466656
VERSION     AJ466656.1  GI:21182612
KEYWORDS    EST.
SOURCE      Hordeum vulgare
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Saren,A.-M., Tanskanen,J., Paulin,L. and Schulman,A.H.
  TITLE     Barley EST's
  JOURNAL   Unpublished
COMMENT     Contact: Schulman AH
            Institute of Biotechnology
            University of Helsinki
            P.O.Box 56 (Viikinkaari 6A), University of Helsinki FIN-00014,
            Finland.
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Hordeum vulgare"
                     /mol_type="mRNA"
                     /db_xref="taxon:4513"
                     /clone="S0000800177G09F1"
                     /tissue_type="Callus"
                     /clone_lib="S00008"
                     /note="Callus K19"
BASE COUNT       97 a    146 c    154 g     83 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 9;  Length 480;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7895 GATGCCGAGGGAAAACTGGAA 7915
              |||||||||||||||||||||
Db          8 GATGCCGAGGGAAAACTGGAA 28


RESULT 39
BH398988/c
LOCUS       BH398988                529 bp    DNA     linear   GSS 11-DEC-2001
DEFINITION  AG-ND-126N2.TR ND-TAM Anopheles gambiae genomic clone AG-ND-126N2,
            genomic survey sequence.
ACCESSION   BH398988
VERSION     BH398988.1  GI:17345204
KEYWORDS    GSS.
SOURCE      Anopheles gambiae (African malaria mosquito)
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 529)
  AUTHORS   Hong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J., Ren
            ,C., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B., Gardner,M.J.
            and Collins,F.H.
  TITLE     Construction of a BAC library and generation of BAC end
            sequence-tagged connectors for genome sequencing of the African
            malaria mosquito Anopheles gambiae
  JOURNAL   Mol. Genet. Genomics 268 (6), 720-728 (2003)
  MEDLINE   22542063
   PUBMED   12655398
COMMENT     Other_GSSs: AG-ND-126N2.TF
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M Universty BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 Rev
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .529
                     /organism="Anopheles gambiae"
                     /mol_type="genomic DNA"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="AG-ND-126N2"
                     /clone_lib="ND-TAM"
                     /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT       84 a    154 c    188 g    103 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 28;  Length 529;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7011 CGCAGCCCGCATCCACCAGGG 7031
              |||||||||||||||||||||
Db         81 CGCAGCCCGCATCCACCAGGG 61


RESULT 40
BI540646
LOCUS       BI540646                531 bp    mRNA    linear   EST 30-AUG-2001
DEFINITION  453992 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BI540646
VERSION     BI540646.1  GI:15381758
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 531)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
```

```
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
   PUBMED   11282978
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 117  row: M  column: 23
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .531
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 1BOV"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from lymph node, ovary,
                     fat, hypothalamus, and pituitary."
BASE COUNT      179 a    117 c    102 g    133 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 12;  Length 531;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3704 ATAGCCTTTCTTCCCCACAGC 3724
              |||||||||||||||||||||
Db        113 ATAGCCTTTCTTCCCCACAGC 133


RESULT 41
AQ680406/c
LOCUS       AQ680406                537 bp    DNA     linear   GSS 25-JUN-1999
DEFINITION  HS_5496_B1_F12_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1072 Col=23 Row=L, genomic survey sequence.
ACCESSION   AQ680406
VERSION     AQ680406.1  GI:5229291
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 537)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1072  row: L  column: 23
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 537.
FEATURES             Location/Qualifiers
     source          1. .537
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=1072 Col=23 Row=L"
                     /sex="male"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      192 a     90 c    114 g    136 t      5 others
ORIGIN

  Query Match             0.2%;  Score 21;  DB 28;  Length 537;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7675 CGTGAGAATTCTCACCGAGAG 7695
              |||||||||||||||||||||
Db        438 CGTGAGAATTCTCACCGAGAG 418


RESULT 42
AZ386696
LOCUS       AZ386696                551 bp    DNA     linear   GSS 02-OCT-2000
DEFINITION  1M0145L17R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0145L17 R, genomic survey sequence.
ACCESSION   AZ386696
VERSION     AZ386696.1  GI:10500396
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 551)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0145  row: L  column: 17
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 551.
FEATURES             Location/Qualifiers
     source          1. .551
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0145L17"
```

```
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      168 a    107 c     96 g    180 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 28;  Length 551;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7940 CTTGTTTTTCAACCCATCTTT 7960
             |||||||||||||||||||||
Db        135 CTTGTTTTTCAACCCATCTTT 155


RESULT 43
BU338432/c
LOCUS       BU338432                 553 bp    mRNA    linear   EST 28-NOV-2002
DEFINITION  603406611F1 CSEQCHN66 Gallus gallus cDNA clone ChEST314e4 5', mRNA
            sequence.
ACCESSION   BU338432
VERSION     BU338432.1  GI:25846433
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 553)
  AUTHORS   Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
  TITLE     A Comprehensive Collection of Chicken cDNAs
  JOURNAL   Curr. Biol. 12 (22), 1965-1969 (2002)
  MEDLINE   22335534
   PUBMED   12445392
COMMENT     Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology (UMIST
            )
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .553
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="White Leghorn, Hisex"
                     /db_xref="taxon:9031"
                     /clone="ChEST314e4"
                     /dev_stage="16 day embryo"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHN66"
                     /note="Organ: brain; Vector: pBluescript II KS(+); Site_1:
                     EcoRI; Site_2: Not1; This normalized library was
                     constructed from 1 million independent clones. cDNA
                     synthesis was initiated using an oligo(dT) primer, using
                     methylated C in the first strand synthesis reaction.
                     Following this first strand reaction, double-stranded cDNA
                     was blunted, ligated to NotI adapters, digested with EcoRI
                     , size-selected, and cloned into the NotI and EcoRI
                     compatible sites of a custom modified MCS of the
                     pBluescript (KS+) vector. The library was normalized in 2
                     rounds using conditions adapted from Soares et al., PNAS
                     (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
                     (1996): 791, except that a significantly longer
                     reannealing hybridization was used."
BASE COUNT       69 a    184 c    196 g    104 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 13;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8692 GTCCCGCATGGCGGCCGGCTC 8712
              |||||||||||||||||||||
Db        133 GTCCCGCATGGCGGCCGGCTC 113


RESULT 44
BZ359326
LOCUS       BZ359326                 570 bp    DNA     linear   GSS 18-NOV-2002
DEFINITION  id72c04.b1 WGS-ZmaysF (JM107 adapted methyl filtered) Zea mays
            genomic clone id72c04 5', genomic survey sequence.
ACCESSION   BZ359326
VERSION     BZ359326.1  GI:25058995
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Rabinowicz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N.,
            Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,
            Zutavern,T., McCombie,W.R. and Martienssen,R.A.
  TITLE     Genomic shotgun sequences from Zea mays (methyl-filtered)
  JOURNAL   Unpublished
COMMENT     Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mccombie@cshl.org
            Plate: id72  row: c  column: 04
            Seq primer: -21M13UnivFwd
            Class: shotgun
            High quality sequence stop: 570.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="id72c04"
                     /lab_host="JM107 or DH5a"
                     /clone_lib="WGS-ZmaysF (JM107 adapted methyl filtered)"
                     /note="Organ: immature ears; Site_1: Xba I; Site_2: Xba I;
                     The vector was digested with XbaI and one nucleotide was
                     added by  fill in in the recessive 3' end. The genomic DNA
                     was nebulized, end  repaired, adaptor ligated and size
                     fractionated using sephadex. The  resulting fragments were
                     between 0.8 and 3 kb and were cloned into  the vector
                     (.x/y reads in M13mp19, .b/g reads in pUC19). The same
                     ligation was transformed in either JM107 or DH5a. "
BASE COUNT      138 a    125 c    141 g    166 t
```

```
ORIGIN

  Query Match             0.2%;  Score 21;  DB 29;  Length 570;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1962 TTTGCCACTCAGTGATGACCT 1982
              |||||||||||||||||||||
Db          3 TTTGCCACTCAGTGATGACCT 23


RESULT 45
BQ992662
LOCUS       BQ992662                599 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  QGF26E04.yg.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
            QGF26E04, mRNA sequence.
ACCESSION   BQ992662
VERSION     BQ992662.1  GI:22412197
KEYWORDS    EST.
SOURCE      Lactuca sativa
  ORGANISM  Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
            Cichorieae; Lactuca.
REFERENCE   1  (bases 1 to 599)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J., Ellison
            ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
            Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QG_CA_Contig4641, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QGF26  row: E  column: 04.
FEATURES             Location/Qualifiers
     source          1. .599
                     /organism="Lactuca sativa"
                     /mol_type="mRNA"
                     /cultivar="L.serriola"
                     /db_xref="taxon:4236"
                     /clone="QGF26E04"
                     /lab_host="E.coli"
                     /clone_lib="QG_EFGHJ lettuce serriola"
                     /note="Vector: pBRcDNASfiAB; The library was constructed
                     from 10 different sources of RNA from a single genotype.
                     Separate cDNAs were generated using primers that
                     incorporated unique 5' and 3' tags to distinguish each
                     source of RNA. cDNAs were then pooled, size-fractionated,
                     directionally cloned into a custom medium-copy vector and
                     transformations made with four size classes to minimize
                     size bias. Details of each source of RNA and library
                     construction can be obtained at http://cgpdb.ucdavis.edu/
                     TAG_LIB=QG_EFGHJ lettuce serriola
                     TAG_TISSUE=flowers post-fertilized
                     TAG_SEQ=TGCCATCGGG"
BASE COUNT      230 a     97 c    156 g    116 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 13;  Length 599;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4931 AGAAATGACTACACAAACACA 4951
              |||||||||||||||||||||
Db         76 AGAAATGACTACACAAACACA 96


RESULT 46
BZ294610/c
LOCUS       BZ294610                601 bp    DNA     linear   GSS 31-OCT-2002
DEFINITION  CG1002.f1 Candida glabrata Random Genomic Library Candida glabrata
            genomic clone CG1002, genomic survey sequence.
ACCESSION   BZ294610
VERSION     BZ294610.1  GI:24435346
KEYWORDS    GSS.
SOURCE      Candida glabrata
  ORGANISM  Candida glabrata
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 601)
  AUTHORS   Wong,S., Fares,M.A., Zimmermann,W., Butler,G. and Wolfe,K.H.
  TITLE     Evidence from comparative genomics for a complete sexual cycle in
            the 'asexual' pathogenic yeast Candida glabrata
  JOURNAL   Genome Biol. 4 (2), R10 (2003)
  MEDLINE   22508158
   PUBMED   12620120
COMMENT     Contact: Wong S
            Department of Genetics, Smurfit Institute
            Trinity College Dublin
            Dublin 2, Ireland
            Tel: 353 1 6082319
            Fax: 353 1 6798558
            Email: swong@tcd.ie
            Class: plasmid ends.
FEATURES             Location/Qualifiers
     source          1. .601
                     /organism="Candida glabrata"
                     /mol_type="genomic DNA"
                     /strain="CBS 138"
                     /db_xref="taxon:5478"
                     /clone="CG1002"
                     /clone_lib="Candida glabrata Random Genomic Library"
BASE COUNT      210 a    106 c    109 g    176 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 29;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4458 AAGCTCTTCTCTGGGGACACT 4478
              |||||||||||||||||||||
Db        563 AAGCTCTTCTCTGGGGACACT 543


RESULT 47
CB977199/c
LOCUS       CB977199                617 bp    mRNA    linear   EST 01-MAY-2003
DEFINITION  CAB40003_IVa_Ra_D11 Cabernet Sauvignon Berry - CAB4 Vitis vinifera
            cDNA clone CAB40003_IVa_Ra_D11 3', mRNA sequence.
ACCESSION   CB977199
VERSION     CB977199.1  GI:30300405
KEYWORDS    EST.
SOURCE      Vitis vinifera
  ORGANISM  Vitis vinifera
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; Vitaceae; Vitis.
REFERENCE   1  (bases 1 to 617)
  AUTHORS   Goes da Silva,F., Iandolino,A., Lim,H., Baek,J., Jones,K. and Cook
            ,D.
  TITLE     Expressed sequence tags from Vitis vinifera 'Cabernet sauvignon'
            berries at various developmental stages
  JOURNAL   Unpublished
COMMENT     Contact: Douglas Cook, PhD
            CAES Genome Facility
            UC Davis, Plant Pathology
            One Shields Ave, Davis, CA 95616, USA
```

```
            Tel: 530 754 6561
            Fax: 530 754 6617
            Email: drcook@ucdavis.edu
            Seq primer: GCCAAACGAATGGTCTAG.
FEATURES             Location/Qualifiers
     source          1. .617
                     /organism="Vitis vinifera"
                     /mol_type="mRNA"
                     /cultivar="Cabernet Sauvignon"
                     /db_xref="taxon:29760"
                     /clone="CAB40003_IVa_Ra_D11"
                     /sex="Hermaphrodite"
                     /dev_stage="Berry on stage II, 9 mm"
                     /lab_host="DH5alpha"
                     /clone_lib="Cabernet Sauvignon Berry - CAB4"
                     /note="Organ: Berry; Vector: pDNR; Site_1: SfiI; Site_2:
                     SfiI; CAB4 is a cDNA library of Vitis vinifera cv.
                     'Cabernet Sauvignon' Clone 8 berries. Sampled berries were
                     collected from field-grown vines during stage II of berry
                     growth (berries were green and hard) at approximately 60
                     days after full bloom.The average berry size was 9
                     milimeters. Sampled vines were located at the University
                     of California, Davis, Experimental Vineyard. cDNAs were
                     made by oligo-dT priming and directionally cloned. 5'and
                     3' adaptors were used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Library was
                     constructed using the Clontech Creator SMART kit and
                     size-selected to contain the 0.5-3 kb size fraction."
BASE COUNT      191 a    129 c     94 g    203 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 14;  Length 617;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8865 AGGGCTTATTCGATTTAATCG 8885
              |||||||||||||||||||||
Db        120 AGGGCTTATTCGATTTAATCG 100


RESULT 48
DR18A19S
LOCUS       DR18A19S                620 bp    DNA     linear   GSS 27-NOV-2002
DEFINITION  Danio rerio genomic clone DKEY-18A19, genomic survey sequence.
ACCESSION   AL745916
VERSION     AL745916.1  GI:21343528
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 620)
  AUTHORS   Humphray,S.J., Huckle,E. and Hunt,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2002) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 18A19. 18A19 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene.
            Further details: http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .620
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-18A19"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
BASE COUNT      220 a    111 c    109 g    180 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 29;  Length 620;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1367 AAACAAAAAAATGCATATGTG 1387
              |||||||||||||||||||||
Db        180 AAACAAAAAAATGCATATGTG 200


RESULT 49
CC372615/c
LOCUS       CC372615                621 bp    DNA     linear   GSS 19-MAY-2003
DEFINITION  PUHGJ24TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa445C24,
            genomic survey sequence.
ACCESSION   CC372615
VERSION     CC372615.1  GI:30846232
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 621)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
            ,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished
COMMENT     Other_GSSs: PUHGJ24TB
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .621
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa445C24"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
BASE COUNT      197 a     95 c    106 g    223 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 29;  Length 621;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4868 TCAAATAATATAACACAAAAT 4888
              |||||||||||||||||||||
Db        241 TCAAATAATATAACACAAAAT 221


RESULT 50
BG624554/c
LOCUS       BG624554                649 bp    mRNA    linear   EST 01-JUL-2001
DEFINITION  CvHEM-037 CvHEM Crassostrea virginica cDNA similar to c-jun protein
            , mRNA sequence.
ACCESSION   BG624554
VERSION     BG624554.1  GI:14580984
KEYWORDS    EST.
SOURCE      Crassostrea virginica (eastern oyster)
  ORGANISM  Crassostrea virginica
            Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Ostreoida;
            Ostreoidea; Ostreidae; Crassostrea.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Gross,P.S. and Bartlett,T.C.
```

```
  TITLE .   Crassostrea virginica Hemocyte EST Library (CvHEM)
  JOURNAL   Unpublished
COMMENT     Contact: Jenny MJ
            Marine Biomedicine and Environmental Sciences
            Medical University of South Carolina
            221 Fort Johnson Road, Charleston, SC 29412, USA
            Tel: 843 876 5058
            Fax: 843 762 5530
            Email: jennymj@musc.edu
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .649
                     /organism="Crassostrea virginica"
                     /mol_type="mRNA"
                     /db_xref="taxon:6565"
                     /tissue_type="Hemocyte"
                     /clone_lib="CvHEM"
                     /note="Vector: pTriplEx2; Site_1: Sfi I; Site_2: Sfi I"
BASE COUNT      209 a    141 c    130 g    158 t     11 others
ORIGIN

  Query Match             0.2%;  Score 21;  DB 10;  Length 649;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2969 ATCGTCTTCGCTGTCTGATTC 2989
              |||||||||||||||||||||
Db        508 ATCGTCTTCGCTGTCTGATTC 488


RESULT 51
BF294496
LOCUS       BF294496                653 bp    mRNA    linear   EST 04-MAY-2001
DEFINITION  005PbG01 Pb cDNA #17, Tommaso Pace, Marta Ponzi, and Clara Frontali
            Plasmodium berghei cDNA 5', mRNA sequence.
ACCESSION   BF294496
VERSION     BF294496.1  GI:13943578
KEYWORDS    EST.
SOURCE      Plasmodium berghei
  ORGANISM  Plasmodium berghei
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 653)
  AUTHORS   Carlton,J.M.-R. and Dame,J.B.
  TITLE     The Plasmodium vivax and P. berghei gene sequence tag projects
  JOURNAL   Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)
COMMENT     Contact: Dame JB
            Department of Pathobiology, College of Veterinary Medicine
            University of Florida
            2015 SW 23rd Avenue, Bldg 1017, Gainseville, FL 32611-0880, USA
            Tel: 352 392 4700
            Fax: 352 392 9704
            Email: damej@mail.vetmed.ufl.edu
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .653
                     /organism="Plasmodium berghei"
                     /mol_type="mRNA"
                     /strain="ANKA clone HP (gametocyte producer)"
                     /db_xref="taxon:5821"
                     /dev_stage="asynchronous blood stage"
                     /lab_host="Wistar rats"
                     /clone_lib="Pb cDNA #17, Tommaso Pace, Marta Ponzi, and
                     Clara Frontali"
                     /note="Vector: pBluescript II vector DNA, excised from
                     Lamda ZAP II.; Site_1: EcoRI; Site_2: XhoI; Total RNA was
                     extracted from asynchronous blood stage forms of the
                     cloned ANKA isolate of P. berghei, grown in Wistar rats to
                     30% parasitemia and 2-5% gametocytemia. Contaminating host
                     white cells  had previously been removed and final host
                     cell contamination estimated to be approximately 5%.
                     PolyA+ RNA was extracted and reverse transcribed  using an
                     oligo dT-XhoI primer (Lambda ZAP II cDNA cloning kit,
                     Stratagene). Second strand cDNA was made following the
                     manufacturer's  protocol. EcoRI adaptors were ligated to
                     the cDNA, and fragments were  ligated into EcoRI/XhoI
                     digested vector."
BASE COUNT      251 a     73 c     62 g    267 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 10;  Length 653;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4713 AATTAAAAAAAATCCACATGT 4733
              |||||||||||||||||||||
Db        604 AATTAAAAAAAATCCACATGT 624


RESULT 52
BG418750
LOCUS       BG418750                656 bp    mRNA    linear   EST 23-OCT-2001
DEFINITION  HVSMEk0024E11f Hordeum vulgare testa/pericarp EST library
            HVcDNA0013 (normal) Hordeum vulgare subsp. vulgare cDNA clone
            HVSMEk0024E11f, mRNA sequence.
ACCESSION   BG418750
VERSION     BG418750.2  GI:16333810
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 656)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Kannangara,G., von
            Wetstein,D., Akhunov,E., Chin,A., Choi,D.W., Fenton,R.D., Kianian
            ,P., Otto,C., Simons,K., Zhang,D., Begum,D., Frisch,D., Yu,Y.,
            Henry,D., Palmer,M., Rambo,T., Simmons,J., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Morex testa/pericarp cDNA library
  JOURNAL   Unpublished
COMMENT     On Mar 13, 2001 this sequence version replaced gi:13324301.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 457
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 650.
FEATURES             Location/Qualifiers
     source          1. .656
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Morex"
                     /db_xref="taxon:112509"
                     /clone="HVSMEk0024E11f"
                     /tissue_type="testa/pericarp"
                     /lab_host="TJC121"
                     /clone_lib="Hordeum vulgare testa/pericarp EST library
                     HVcDNA0013 (normal)"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1;
                     Plants were raised from seeds in a Controlled Environments
                     growth chamber maintained in continuous light at 18oC, and
                     testa and pericarp were dissected from developing kernels
                     at Washington State University, Pullman, WA (Kannangara,
                     von Wetstein). Total RNA was prepared, poly(A) RNA was
                     purified, one cDNA library was made, and 1 million pfu
                     were in vivo excised to give pBluescript SK(-) cDNA
                     phagemids in the TJ Close lab at the University of
                     California, Riverside (Akhunov, Chin, Choi, Close, Fenton,
                     Kianian, Otto, Simons, Zhang). Phagemids were plated and
                     picked at the Clemson University Genomics Institute (CUGI)
                     (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
                     preparations, DNA sequencing and sequence analysis were
```

performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates , Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT 145 a 185 c 192 g 134 t
ORIGIN

Query Match 0.2%; Score 21; DB 10; Length 656;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7895 GATGCCGAGGGAAAACTGGAA 7915
||||||||||||||||||||
Db 27 GATGCCGAGGGAAAACTGGAA 47

RESULT 53
BZ002130/c
LOCUS BZ002130 684 bp DNA linear GSS 07-OCT-2002
DEFINITION oee36b07.b1 B.oleracea002 Brassica oleracea genomic, genomic survey sequence.
ACCESSION BZ002130
VERSION BZ002130.1 GI:23550388
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 684)
AUTHORS Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash ,W., Rabinowicz,P.D. and Wilson,R.K.
TITLE Whole genome shotgun reads from Brassica oleracea
JOURNAL Unpublished
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Plate: oee36 row: b column: 07
Seq primer: -21UPpOT forward
Class: shotgun
High quality sequence start: 90
High quality sequence stop: 551.
FEATURES Location/Qualifiers
source 1. .684
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pOTw13; Whole genome shotgun library from flowering buds. DNA was purified from a crude nuclear prep using Brassica oleracea TO1000DH3 buds provided by Thosmas Osborn at the University of Wisconsin. Genomic DNA was provided by Pablo Rabinowicz (CSHL) and the shotgun library prepared at Washington University Genome Sequencing Center."
BASE COUNT 193 a 153 c 135 g 203 t
ORIGIN

Query Match 0.2%; Score 21; DB 28; Length 684;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6703 AAAAAAACAAAGATTTAAAAA 6723
||||||||||||||||||||
Db 312 AAAAAAACAAAGATTTAAAAA 292

RESULT 54
AZ278564/c
LOCUS AZ278564 686 bp DNA linear GSS 26-JUL-2000
DEFINITION RPCI-23-119F19.TV RPCI-23 Mus musculus genomic clone RPCI-23-119F19 , genomic survey sequence.
ACCESSION AZ278564
VERSION AZ278564.1 GI:9497466
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 686)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished
COMMENT Other_GSSs: RPCI-23-119F19.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 119 row: F column: 19
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .686
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-119F19"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). "
BASE COUNT 236 a 117 c 106 g 227 t
ORIGIN

Query Match 0.2%; Score 21; DB 28; Length 686;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3921 CTATCTCTTTAAAAATATTTT 3941
||||||||||||||||||||
Db 574 CTATCTCTTTAAAAATATTTT 554

RESULT 55
BZ964704/c
LOCUS BZ964704 693 bp DNA linear GSS 25-MAR-2003
DEFINITION PUGGI13TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa378C01, genomic survey sequence.
ACCESSION BZ964704
VERSION BZ964704.1 GI:29179965

```
KEYWORDS    G6S.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 693)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
            ,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished
COMMENT     Other_GSSs: PUGGI13TB
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .693
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa378C01"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
BASE COUNT      225 a    107 c    110 g    251 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 29;  Length 693;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4868 TCAAATAATATAACACAAAAT 4888
              |||||||||||||||||||||
Db        203 TCAAATAATATAACACAAAAT 183


RESULT 56
BQ988501
LOCUS       BQ988501                695 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  QGF15B13.yg.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
            QGF15B13, mRNA sequence.
ACCESSION   BQ988501
VERSION     BQ988501.1  GI:22408036
KEYWORDS    EST.
SOURCE      Lactuca sativa
  ORGANISM  Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
            Cichorieae; Lactuca.
REFERENCE   1  (bases 1 to 695)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J., Ellison
            ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
            Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QG_CA_Contig4641, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QGF15  row: B  column: 13.
FEATURES             Location/Qualifiers
     source          1. .695
                     /organism="Lactuca sativa"
                     /mol_type="mRNA"
                     /cultivar="L.serriola"
                     /db_xref="taxon:4236"
                     /clone="QGF15B13"
                     /lab_host="E.coli"
                     /clone_lib="QG_EFGHJ lettuce serriola"
                     /note="Vector: pBRcDNASfiAB; The library was constructed
                     from 10 different sources of RNA from a single genotype.
                     Separate cDNAs were generated using primers that
                     incorporated unique 5' and 3' tags to distinguish each
                     source of RNA. cDNAs were then pooled, size-fractionated,
                     directionally cloned into a custom medium-copy vector and
                     transformations made with four size classes to minimize
                     size bias. Details of each source of RNA and library
                     construction can be obtained at http://cgpdb.ucdavis.edu/
                     TAG_LIB=QG_EFGHJ lettuce serriola
                     TAG_TISSUE=leaves dark grow
                     TAG_SEQ=GCTAGTCGGG"
BASE COUNT      275 a     96 c    179 g    145 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 13;  Length 695;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4931 AGAAATGACTACACAAACACA 4951
              |||||||||||||||||||||
Db        120 AGAAATGACTACACAAACACA 140


RESULT 57
BZ067843
LOCUS       BZ067843                705 bp    DNA     linear   GSS 10-OCT-2002
DEFINITION  lkg05f07.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
            sequence.
ACCESSION   BZ067843
VERSION     BZ067843.1  GI:23683876
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 705)
  AUTHORS   Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash
            ,W., Rabinowicz,P.D. and Wilson,R.K.
  TITLE     Whole genome shotgun reads from Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: lkg05  row: f  column: 07
            Seq primer: -21UPpOT forward
            Class: shotgun
            High quality sequence stop: 265.
FEATURES             Location/Qualifiers
     source          1. .705
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3712"
                     /clone_lib="B.oleracea002"
                     /note="Vector: pOTw13; Whole genome shotgun library from
                     flowering buds. DNA was purified from a crude nuclear
                     prep using Brassica oleracea TO1000DH3 buds provided by
                     Thosmas Osborn at the University of Wisconsin. Genomic
                     DNA was provided by Pablo Rabinowicz (CSHL) and the
                     shotgun library prepared at Washington University Genome
                     Sequencing Center."
```

```
BASE COUNT      238 a    128 c    160 g    175 t      4 others
ORIGIN

  Query Match             0.2%;  Score 21;  DB 28;  Length 705;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6775 AAAAAATTTTTTAACCAAACT 6795
              |||||||||||||||||||||
Db        431 AAAAAATTTTTTAACCAAACT 451


RESULT 58
BQ993877
LOCUS       BQ993877                 708 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  QGF5L10.yg.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
            QGF5L10, mRNA sequence.
ACCESSION   BQ993877
VERSION     BQ993877.1  GI:22413412
KEYWORDS    EST.
SOURCE      Lactuca sativa
  ORGANISM  Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
            Cichorieae; Lactuca.
REFERENCE   1  (bases 1 to 708)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J., Ellison
            ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
            Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QG_CA_Contig4641, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QGF5  row: L  column: 10.
FEATURES             Location/Qualifiers
     source          1. .708
                     /organism="Lactuca sativa"
                     /mol_type="mRNA"
                     /cultivar="L.serriola"
                     /db_xref="taxon:4236"
                     /clone="QGF5L10"
                     /lab_host="E.coli"
                     /clone_lib="QG_EFGHJ lettuce serriola"
                     /note="Vector: pBRcDNASfiAB; The library was constructed
                     from 10 different sources of RNA from a single genotype.
                     Separate cDNAs were generated using primers that
                     incorporated unique 5' and 3' tags to distinguish each
                     source of RNA. cDNAs were then pooled, size-fractionated,
                     directionally cloned into a custom medium-copy vector and
                     transformations made with four size classes to minimize
                     size bias. Details of each source of RNA and library
                     construction can be obtained at http://cgpdb.ucdavis.edu/
                     TAG_LIB=QG_EFGHJ lettuce serriola
                     TAG_TISSUE=flowers post-fertilized
                     TAG_SEQ=TGCCATCGGG"
BASE COUNT      249 a    132 c    163 g    163 t      1 others
ORIGIN

  Query Match             0.2%;  Score 21;  DB 13;  Length 708;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4931 AGAAATGACTACACAAACACA 4951
              |||||||||||||||||||||
Db        272 AGAAATGACTACACAAACACA 292


RESULT 59
BH592832/c
LOCUS       BH592832                 721 bp    DNA     linear   GSS 15-DEC-2001
DEFINITION  BOHFK69TR BOHF Brassica oleracea genomic clone BOHFK69, genomic
            survey sequence.
ACCESSION   BH592832
VERSION     BH592832.1  GI:17845284
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 721)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOHFK69TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .721
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHFK69"
                     /clone_lib="BOHF"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      280 a     94 c    117 g    230 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 28;  Length 721;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        527 TATCCAACGCGAATTTGAAA 547
              |||||||||||||||||||||
Db        276 TATCCAACGCGAATTTGAAA 256


RESULT 60
BQ763978
LOCUS       BQ763978                 724 bp    mRNA    linear   EST 26-JUL-2002
DEFINITION  EBro03_SQ005_N22_R root, 3 week, waterlogged, cv Optic, EBro03
            Hordeum vulgare subsp. vulgare cDNA clone EBro03_SQ005_N22 5', mRNA
            sequence.
ACCESSION   BQ763978
VERSION     BQ763978.1  GI:21972450
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 724)
  AUTHORS   Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L.,
            Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
  TITLE     Development of Barley Transcriptome Resources
  JOURNAL   Unpublished
COMMENT     Contact: Waugh R, Marshall DF
```

```
          Genome Dynamics/Computational Biology
          Scottish Crop Research Institute
          Invergowrie, Dundee, DD2 5DA, Scotland, UK
          Tel: 00 44 1382 562731
          Fax: 00 44 1382 562426
          Email: est@scri.sari.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .724
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Optic"
                     /db_xref="taxon:112509"
                     /clone="EBro03_SQ005_N22"
                     /tissue_type="root"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="root, 3 week, waterlogged, cv Optic, EBro03"
                     /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
                     Non-normalised library, directionally cloned into pSPORT1.
                     Derived from roots of 3 week old waterlogged barley
                     plants.  Developed as part of the barley transcriptome
                     resources of BBSRC/SEERAD funded cereal IGF (Investigating
                     Gene Function) project."
BASE COUNT      131 a    220 c    218 g    155 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 13;  Length 724;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7895 GATGCCGAGGGAAAACTGGAA 7915
             |||||||||||||||||||||
Db        369 GATGCCGAGGGAAAACTGGAA 389


RESULT 61
AL965466/c
LOCUS       AL965466                739 bp    mRNA    linear   EST 27-NOV-2002
DEFINITION  AL965466 XGC-gastrula Silurana tropicalis cDNA clone TGas132g03 5',
            mRNA sequence.
ACCESSION   AL965466
VERSION     AL965466.1  GI:25789061
KEYWORDS    EST.
SOURCE      Silurana tropicalis (western clawed frog)
  ORGANISM  Silurana tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Silurana.
REFERENCE   1  (bases 1 to 739)
  AUTHORS   Taylor,R., Ashurst,J.L., Croning,M.D.R., Zorn,A.M. and Rogers,J.
  TITLE     Sanger Xenopus tropicalis EST project 2002
  JOURNAL   Unpublished
COMMENT     Contact: Taylor R
            Sanger Centre
            Hinxton, Cambridgeshire, CB10 1SA, UK
            Email: trop@sanger.ac.uk
            Sanger Xenopus tropicalis EST project 2001
            TROPICALIS_SEQUENCE_ID: TGas132g03.plkSP6
            Sequencing primer: SP6
            This sequence is from a Xenopus Gene Collection (XGC) library
            constructed by Aaron M. Zorn.
FEATURES             Location/Qualifiers
     source          1. .739
                     /organism="Silurana tropicalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8364"
                     /clone="TGas132g03"
                     /dev_stage="gastrula (stages 10.5-13 mixed)"
                     /lab_host="Escherichia coli XL1-blue"
                     /clone_lib="XGC-gastrula"
                     /note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
                     was oligo dT primed from 5ug of poly A+ RNA from stages
                     10-13 gastrulae.  EcoRI-NotI cut cDNA was then ligated
                     into pCS107 with EcoRI at the 5' end and NotI at the 3'
                     end."
BASE COUNT      221 a    130 c    139 g    249 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 9;  Length 739;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3929 TTAAAAATATTTTTGAGGCTT 3949
             |||||||||||||||||||||
Db        394 TTAAAAATATTTTTGAGGCTT 374


RESULT 62
BG596235
LOCUS       BG596235                757 bp    mRNA    linear   EST 07-MAR-2003
DEFINITION  EST494913 cSTS Solanum tuberosum cDNA clone cSTS13J13 5' sequence,
            mRNA sequence.
ACCESSION   BG596235
VERSION     BG596235.1  GI:13614375
KEYWORDS    EST.
SOURCE      Solanum tuberosum (potato)
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 757)
  AUTHORS   van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chiemingo,A.,
            Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
  TITLE     Generations of ESTs from sprouting potato eyes
  JOURNAL   Unpublished
COMMENT     Contact: Robin Buell
            The Institute for Genomic Research
            9712 Medical Center Dr, Rockville, MD 20850, USA
            Email: potato-array@tigr.org
            This clone can be obtained from the University of Arizona Genomics
            Institute. Orders can be made through URL:
            http://genome.arizona.edu/orders/
            Seq primer: M13F-R.
FEATURES             Location/Qualifiers
     source          1. .757
                     /organism="Solanum tuberosum"
                     /mol_type="mRNA"
                     /cultivar="Kennebec"
                     /db_xref="taxon:4113"
                     /clone="cSTS13J13"
                     /tissue_type="sprouting eyes from tubers"
                     /dev_stage="12-14 weeks post harvest"
                     /lab_host="SOLR"
                     /clone_lib="cSTS"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
                     taken from tubers. The tubers were incubated at 26C in the
                     dark for 2-3 weeks prior to sprouting. The eyes were
                     frozen in liquid nitrogen immediately upon removal from
                     tubers."
BASE COUNT      233 a    123 c    149 g    252 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 10;  Length 757;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7281 GATAGTGGAACTTGTACATAT 7301
             |||||||||||||||||||||
Db        376 GATAGTGGAACTTGTACATAT 396


RESULT 63
CB976925/c
LOCUS       CB976925                770 bp    mRNA    linear   EST 01-MAY-2003
DEFINITION  CAB40003_IIIa_Ra_E03 Cabernet Sauvignon Berry - CAB4 Vitis vinifera
```

```
            cDNA clone CAB40003_IIIa_Ra_E03 3', mRNA sequence.
ACCESSION   CB976925
VERSION     CB976925.1  GI:30300131
KEYWORDS    EST.
SOURCE      Vitis vinifera
  ORGANISM  Vitis vinifera
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; Vitaceae; Vitis.
REFERENCE   1  (bases 1 to 770)
  AUTHORS   Goes da Silva,F., Iandolino,A., Lim,H., Baek,J., Jones,K. and Cook
            ,D.
  TITLE     Expressed sequence tags from Vitis vinifera 'Cabernet sauvignon'
            berries at various developmental stages
  JOURNAL   Unpublished
COMMENT     Contact: Douglas Cook, PhD
            CAES Genome Facility
            UC Davis, Plant Pathology
            One Shields Ave, Davis, CA 95616, USA
            Tel: 530 754 6561
            Fax: 530 754 6617
            Email: drcook@ucdavis.edu
            Seq primer: GCCAAACGAATGGTCTAG.
FEATURES             Location/Qualifiers
     source          1. .770
                     /organism="Vitis vinifera"
                     /mol_type="mRNA"
                     /cultivar="Cabernet Sauvignon"
                     /db_xref="taxon:29760"
                     /clone="CAB40003_IIIa_Ra_E03"
                     /sex="Hermaphrodite"
                     /dev_stage="Berry on stage II, 9 mm"
                     /lab_host="DH5alpha"
                     /clone_lib="Cabernet Sauvignon Berry - CAB4"
                     /note="Organ: Berry; Vector: pDNR; Site_1: SfiI; Site_2:
                     SfiI; CAB4 is a cDNA library of Vitis vinifera cv.
                     'Cabernet Sauvignon' Clone 8 berries. Sampled berries were
                     collected from field-grown vines during stage II of berry
                     growth (berries were green and hard) at approximately 60
                     days after full bloom.The average berry size was 9
                     milimeters. Sampled vines were located at the University
                     of California, Davis, Experimental Vineyard. cDNAs were
                     made by oligo-dT priming and directionally cloned. 5'and
                     3' adaptors were used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Library was
                     constructed using the Clontech Creator SMART kit and
                     size-selected to contain the 0.5-3 kb size fraction."
BASE COUNT      256 a    118 c    117 g    279 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 14;  Length 770;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8865 AGGGCTTATTCGATTTAATCG 8885
              |||||||||||||||||||||
Db        633 AGGGCTTATTCGATTTAATCG 613


RESULT 64
AJ514001
LOCUS       AJ514001                773 bp    mRNA    linear   EST 14-NOV-2002
DEFINITION  AJ514001 Gossypium hirsutum ovule first day the flower opens
            Gossypium hirsutum cDNA clone suo57d07r3, mRNA sequence.
ACCESSION   AJ514001
VERSION     AJ514001.1  GI:24977484
KEYWORDS    EST.
SOURCE      Gossypium hirsutum (upland cotton)
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 773)
  AUTHORS   Suo,J., Liang,X. and Xue,Y.
  TITLE     The construction of a cDNA library from the ovule of Gossypium
            hirsutum and its EST analysis
  JOURNAL   Unpublished
COMMENT     Contact: Suo J
            Centre of Developmental Biology
            The Institute of Genetics and Developmental Biology CAS
            Nanyitiao 3, Zhongguancun, Haidian, Beijing, 100080, China.
FEATURES             Location/Qualifiers
     source          1. .773
                     /organism="Gossypium hirsutum"
                     /mol_type="mRNA"
                     /cultivar="Zhongmian12"
                     /db_xref="taxon:3635"
                     /clone="suo57d07r3"
                     /tissue_type="ovule"
                     /dev_stage="first day the flower opens"
                     /clone_lib="Gossypium hirsutum ovule first day the flower
                     opens"
BASE COUNT      284 a    192 c    130 g    167 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 9;  Length 773;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4931 AGAAATGACTACACAAACACA 4951
              |||||||||||||||||||||
Db        614 AGAAATGACTACACAAACACA 634


RESULT 65
AG160514/c
LOCUS       AG160514                805 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-026F21.TJ, genomic survey
            sequence.
ACCESSION   AG160514
VERSION     AG160514.1  GI:16690192
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 805)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                 Sequencing: TJ
             LIBRARY
                 Vector    : pBACe3.6
                 R.Site 1  : EcoRI
                 R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .805
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-026F21.TJ"
```

```
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT      349 a    127 c    136 g    193 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 29;  Length 805;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3921 CTATCTCTTTAAAAATATTTT 3941
              |||||||||||||||||||||
Db        413 CTATCTCTTTAAAAATATTTT 393


RESULT 66
BG697867/c
LOCUS       BG697867                 848 bp    mRNA    linear   EST 07-MAY-2001
DEFINITION  602661083F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4804220 5',
            mRNA sequence.
ACCESSION   BG697867
VERSION     BG697867.1  GI:13964563
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 848)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
            Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10701  row: e  column: 21
            High quality sequence stop: 766.
FEATURES             Location/Qualifiers
     source          1. .848
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4804220"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Skn3"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.5kb.  Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      210 a    209 c    213 g    216 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 10;  Length 848;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1975 GATGACCTCAGGAAGAAGATG 1995
              |||||||||||||||||||||
Db        229 GATGACCTCAGGAAGAAGATG 209


RESULT 67
BX222878
LOCUS       BX222878                 850 bp    DNA     linear   GSS 29-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-269P23, genomic survey sequence.
ACCESSION   BX222878
VERSION     BX222878.1  GI:28054764
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 850)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 269P23. 269P23
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .850
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-269P23"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
BASE COUNT      305 a    123 c    130 g    292 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 29;  Length 850;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4712 TAATTAAAAAAAATCCACATG 4732
              |||||||||||||||||||||
Db        201 TAATTAAAAAAAATCCACATG 221


RESULT 68
BG624651/c
LOCUS       BG624651                 914 bp    mRNA    linear   EST 01-JUL-2001
DEFINITION  CvHEM-162 CvHEM Crassostrea virginica cDNA similar to Transcription
            Factor AP-1, mRNA sequence.
ACCESSION   BG624651
VERSION     BG624651.1  GI:14581081
KEYWORDS    EST.
SOURCE      Crassostrea virginica (eastern oyster)
  ORGANISM  Crassostrea virginica
            Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Ostreoida;
            Ostreoidea; Ostreidae; Crassostrea.
REFERENCE   1  (bases 1 to 914)
  AUTHORS   Gross,P.S. and Bartlett,T.C.
  TITLE     Crassostrea virginica Hemocyte EST Library (CvHEM)
  JOURNAL   Unpublished
COMMENT     Contact: Jenny MJ
            Marine Biomedicine and Environmental Sciences
            Medical University of South Carolina
            221 Fort Johnson Road, Charleston, SC 29412, USA
            Tel: 843 876 5058
            Fax: 843 762 5530
            Email: jennymj@musc.edu
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .914
                     /organism="Crassostrea virginica"
                     /mol_type="mRNA"
                     /db_xref="taxon:6565"
                     /tissue_type="Hemocyte"
                     /clone_lib="CvHEM"
                     /note="Vector: pTriplEx2; Site_1: Sfi I; Site_2: Sfi I"
BASE COUNT      293 a    203 c    178 g    214 t     26 others
ORIGIN

  Query Match             0.2%;  Score 21;  DB 10;  Length 914;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       2969 ATCGTCTTCGCTGTCTGATTC 2989
              |||||||||||||||||||||
Db        585 ATCGTCTTCGCTGTCTGATTC 565


RESULT 69
BG243673/c
LOCUS       BG243673                969 bp    mRNA    linear   EST 13-FEB-2001
DEFINITION  602357763F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4486455 5',
            mRNA sequence.
ACCESSION   BG243673
VERSION     BG243673.1  GI:12753488
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 969)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10329  row: m  column: 16
            High quality sequence stop: 664.
FEATURES             Location/Qualifiers
     source          1. .969
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4486455"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="10 months, virgin"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam1"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      284 a    222 c    298 g    165 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 10;  Length 969;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        155 TGGCCTTCGATCCGGCTGAAG 175
              |||||||||||||||||||||
Db        159 TGGCCTTCGATCCGGCTGAAG 139


RESULT 70
BM564204
LOCUS       BM564204                1282 bp    mRNA    linear   EST 20-FEB-2002
DEFINITION  AGENCOURT_6560135 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5742152
            5', mRNA sequence.
ACCESSION   BM564204
VERSION     BM564204.1  GI:18811809
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1282)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12759  row: n  column: 09
            High quality sequence start: 70
            High quality sequence stop: 314.
FEATURES             Location/Qualifiers
     source          1. .1282
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5742152"
                     /tissue_type="medulla"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_119"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source normal medulla from
                     anonymous male age 27.  Library  is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed  upon
                     cloning).  Average insert size 1.3 kb, insert size range
                     0.9-3 kb.  Library is normalized and enriched for
                     full-length clones and  was constructed by C. Gruber
                     (Invitrogen).  Research Genetics  tracking code 013. Note:
                     this is a NIH_MGC Library."
BASE COUNT      323 a    365 c    301 g    293 t
ORIGIN

  Query Match             0.2%;  Score 21;  DB 12;  Length 1282;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        229 GCTGGGAAAAGTACCAGCATA 249
              |||||||||||||||||||||
Db        729 GCTGGGAAAAGTACCAGCATA 749


RESULT 71
CC190145
LOCUS       CC190145                1412 bp    DNA     linear   GSS 08-MAY-2003
DEFINITION  CH261-37F5_RM1.1 CH261 Gallus gallus genomic clone CH261-37F5,
            genomic survey sequence.
ACCESSION   CC190145
VERSION     CC190145.1  GI:30435100
KEYWORDS    GSS.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1412)
  AUTHORS   Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
            Warren,W., Graves,T., Mardis,E. and Wilson,R.
  TITLE     Gallus gallus BAC End Reads
  JOURNAL   Unpublished
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 182000   Std Error: 0.00
            Seq primer: RM1 TACGACTCACTATAGGGAGA
            Class: BAC ends
            High quality sequence start: 255
            High quality sequence stop: 1041.
FEATURES             Location/Qualifiers
```

```
     Source          1. .1412
                     /organism="Gallus gallus"
                     /mol_type="genomic DNA"
                     /strain="Red Jungle Fowl"
                     /db_xref="taxon:9031"
                     /clone="CH261-37F5"
                     /sex="female"
                     /cell_line="UCD001, inbred 256"
                     /clone_lib="CH261"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CH261 Female Chicken library - for library and clone
                     ordering information: http://www.chori.org/bacpac"
BASE COUNT      351 a     281 c     234 g     336 t     210 others
ORIGIN

  Query Match             0.2%;  Score 21;  DB 29;  Length 1412;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        431 AAATTCAGCAACAGGAGCTAT 451
              |||||||||||||||||||||
Db        878 AAATTCAGCAACAGGAGCTAT 898


RESULT 72
BE936464
LOCUS       BE936464                 113 bp    mRNA    linear   EST 02-OCT-2000
DEFINITION  RC1-NT0033-310800-019-a06 NT0033 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE936464
VERSION     BE936464.1  GI:10462540
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 113)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=RC1-NT0033-310
            800-019-a06&t3=2000-08-31&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 32
            High quality sequence stop: 113.
FEATURES             Location/Qualifiers
     source          1. .113
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="NT0033"
                     /note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       47 a     21 c     17 g     28 t
ORIGIN

  Query Match             0.2%;  Score 20;  DB 10;  Length 113;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
              ||||||||||||||||||||
Db         92 GCAAACATCAAATAATATAA 111


RESULT 73
AW419065/c
LOCUS       AW419065                 162 bp    mRNA    linear   EST 09-FEB-2000
DEFINITION  xy46g12.x1 NCI_CGAP_Lu34.1 Homo sapiens cDNA clone IMAGE:2856262
            3', mRNA sequence.
ACCESSION   AW419065
VERSION     AW419065.1  GI:6946997
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Michael Kelley M.D., John Gillespie M.D.,
            Michael Emmert-Buck M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Possible reversed clone: polyT not found
            Seq primer: -40UP from Gibco
            High quality sequence stop: 96.
FEATURES             Location/Qualifiers
     source          1. .162
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2856262"
                     /tissue_type="large cell carcinoma"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Lu34.1"
                     /note="Organ: lung; Vector: pAMP10; cDNA made by oligo-dT
                     priming. Non-directionally cloned into the UDG sites of
                     pAMP10. Size-selected on agarose gel, average insert  size
                     500 bp. Primary library; non-amplified.  cDNA Library
                     Preparation: David B. Krizman, Ph.D (NCI). Reference:
                     Krizman et al. (1996) Cancer Research 56:5380-5383."
BASE COUNT       42 a     24 c     31 g     65 t
ORIGIN

  Query Match             0.2%;  Score 20;  DB 9;  Length 162;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
              ||||||||||||||||||||
Db         56 GCAAACATCAAATAATATAA 37
```

```
RESULT 74
BF369525/c
LOCUS       BF369525                163 bp    mRNA    linear   EST 24-NOV-2000
DEFINITION  CM3-GN0092-230900-368-f06 GN0092 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF369525
VERSION     BF369525.1  GI:11331550
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be.seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM3&t2=CM3-GN0092-
            230900-368-f06&t3=2000-09-23&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 22
            High quality sequence stop: 163.
FEATURES             Location/Qualifiers
     source          1. .163
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="GN0092"
                     /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                     ; Site_2: SmaI; A mini-library was made by cloning
                     products derived from  ORESTES PCR (U.S. Letters Patent
                     application No. 196,716 -  Ludwig Institute for Cancer
                     Research) profiles into  the pUC 18 vector. Reverse
                     transcription of tissue mRNA and cDNA amplification were
                     performed under low stringency conditions."
BASE COUNT       48 a     21 c     30 g     64 t
ORIGIN

  Query Match             0.2%;  Score 20;  DB 10;  Length 163;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
              ||||||||||||||||||||
Db        127 GCAAACATCAAATAATATAA 108


RESULT 75
BE000214
LOCUS       BE000214                188 bp    mRNA    linear   EST 05-JUN-2000
DEFINITION  MR0-BN0070-210400-016-d08 BN0070 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE000214
VERSION     BE000214.1  GI:8260447
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 188)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=MR0-BN0070-210
            400-016-d08&t3=2000-04-21&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 17
            High quality sequence stop: 188.
FEATURES             Location/Qualifiers
     source          1. .188
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="BN0070"
                     /note="Organ: breast_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       68 a     39 c     34 g     47 t
ORIGIN

  Query Match             0.2%;  Score 20;  DB 10;  Length 188;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6704 AAAAAACAAAGATTTAAAAA 6723
              ||||||||||||||||||||
Db         61 AAAAAACAAAGATTTAAAAA 80


RESULT 76
BF327270/c
LOCUS       BF327270                191 bp    mRNA    linear   EST 22-NOV-2000
DEFINITION  MR0-BN0070-250800-034-d10 BN0070 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF327270
VERSION     BF327270.1  GI:11298018
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 191)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
```

```
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR0&t2=MR0-BN0070-
            250800-034-d10&t3=2000-08-25&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 5
            High quality sequence stop: 191.
FEATURES             Location/Qualifiers
     source          1. .191
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="BN0070"
                     /note="Organ: breast_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       47 a     38 c     36 g     69 t      1 others
ORIGIN

  Query Match             0.2%;  Score 20;  DB 10;  Length 191;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6704 AAAAAACAAAGATTTAAAAA 6723
              ||||||||||||||||||||
Db        141 AAAAAACAAAGATTTAAAAA 122


RESULT 77
AZ308718
LOCUS       AZ308718                 202 bp    DNA     linear   GSS 29-SEP-2000
DEFINITION  1M0012D22F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0012D22 F, genomic survey sequence.
ACCESSION   AZ308718
VERSION     AZ308718.1  GI:10348997
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 202)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0012  row: D  column: 22
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 202.
FEATURES             Location/Qualifiers
     source          1. .202
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0012D22"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       50 a     53 c     49 g     50 t
ORIGIN

  Query Match             0.2%;  Score 20;  DB 28;  Length 202;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1696 ACCTCCCTCCAGGTTTTGGT 1715
              ||||||||||||||||||||
Db        123 ACCTCCCTCCAGGTTTTGGT 142


RESULT 78
AW999124
LOCUS       AW999124                 206 bp    mRNA    linear   EST 05-JUN-2000
DEFINITION  MR0-BN0070-010300-003-c11 BN0070 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW999124
VERSION     AW999124.1  GI:8259358
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 206)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
```

```
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=MR0-BN0070-010
            300-003-c11&t3=2000-03-01&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 206.
FEATURES              Location/Qualifiers
     source           1. .206
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /dev_stage="Adult"
                      /clone_lib="BN0070"
                      /note="Organ: breast_normal; Vector: puc18; Site_1: SmaI;
                      Site_2: SmaI; A mini-library was made by cloning products
                      derived from  ORESTES PCR (U.S. Letters Patent application
                      No. 196,716 -  Ludwig Institute for Cancer Research)
                      profiles into  the pUC 18 vector. Reverse transcription of
                      tissue mRNA and cDNA amplification were performed under
                      low stringency conditions."
BASE COUNT       74 a     40 c     38 g     54 t
ORIGIN

  Query Match             0.2%;  Score 20;  DB 10;  Length 206;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6704 AAAAAACAAAGATTTAAAAA 6723
              ||||||||||||||||||||
Db         65 AAAAAACAAAGATTTAAAAA 84


RESULT 79
BB539373
LOCUS       BB539373                209 bp    mRNA    linear   EST 31-JUL-2000
DEFINITION  BB539373 RIKEN full-length enriched, 0 day neonate eyeball Mus
            musculus cDNA clone E130018J09 3' similar to AF093139 Rattus
            norvegicus tip associating protein (TAP) mRNA, mRNA sequence.
ACCESSION   BB539373
VERSION     BB539373.1  GI:9594873
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 209)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu;N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES              Location/Qualifiers
     source           1. .209
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /db_xref="taxon:10090"
                      /clone="E130018J09"
                      /tissue_type="eyeball"
                      /dev_stage="0 day neonate"
                      /lab_host="DH10B"
                      /clone_lib="RIKEN full-length enriched, 0 day neonate
                      eyeball"
                      /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                      prepared and sequenced in Mouse Genome Encyclopedia
                      Project of Genome Exploration Research Group in Riken
                      Genomic Sciences Center and Genome Science Laboratory in
                      RIKEN. Division of Experimental Animal Research in Riken
                      contributed to prepare mouse tissues. 1st strand cDNA was
                      primed with a primer [5'
                      GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
                      prepared by using trehalose thermo-activated reverse
                      transcriptase and subsequently enriched for full-length by
                      cap-trapper. Second strand cDNA was prepared with the
                      primer adapter of sequence [5'
                      GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                      was cleaved with BamHI and XhoI. Vector: a modified
                      pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT       53 a     51 c     48 g     57 t
ORIGIN

  Query Match             0.2%;  Score 20;  DB 10;  Length 209;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6530 AGGTGCTCTGTGTACTTTAC 6549
              ||||||||||||||||||||
Db        148 AGGTGCTCTGTGTACTTTAC 167


RESULT 80
DR23F24S/c
LOCUS       DR23F24S                216 bp    DNA     linear   GSS 21-NOV-2002
DEFINITION  Danio rerio genomic clone DKEY-23F24, genomic survey sequence.
ACCESSION   AL975255
VERSION     AL975255.1  GI:25174401
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 216)
  AUTHORS   Humphray,S.J., Huckle,E. and Hunt,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2002) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 23F24. 23F24 is
```

```
          part of the Daniokey BAC Library created by R. Plasterk and N.V.
          Keygene.
          Further details: http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .216
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-23F24"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
BASE COUNT       88 a     12 c     23 g     90 t      3 others
ORIGIN

  Query Match             0.2%;  Score 20;  DB 29;  Length 216;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6704 AAAAAACAAAGATTTAAAAA 6723
              ||||||||||||||||||||
Db        112 AAAAAACAAAGATTTAAAAA 93


RESULT 81
BI001235/c
LOCUS       BI001235                 220 bp    mRNA    linear   EST 13-JUN-2001
DEFINITION  MR2-HN0036-110101-011-f06 HN0036 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BI001235
VERSION     BI001235.1  GI:14405309
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 220)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR2&t2=MR2-HN0036-
            110101-011-f06&t3=2001-01-11&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 38
            High quality sequence stop: 187.
FEATURES             Location/Qualifiers
     source          1. .220
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="HN0036"
                     /note="Organ: head_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       61 a     65 c     50 g     44 t
ORIGIN

  Query Match             0.2%;  Score 20;  DB 12;  Length 220;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6152 AGTTAGACTCTGTCTCCCTG 6171
              ||||||||||||||||||||
Db         38 AGTTAGACTCTGTCTCCCTG 19


RESULT 82
BQ785339/c
LOCUS       BQ785339                 227 bp    mRNA    linear   EST 26-JUL-2002
DEFINITION  saq76c10.y1 Gm-c1076 Glycine max cDNA clone SOYBEAN CLONE ID:
            Gm-c1076-5516 5', mRNA sequence.
ACCESSION   BQ785339
VERSION     BQ785339.1  GI:21993811
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 227)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact: ccu@resgen.com web site:
            www.resgen.com
            Putative full length read
            vector to vector length is 228
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .227
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /db_xref="taxon:3847"
                     /clone="SOYBEAN CLONE ID: Gm-c1076-5516"
                     /tissue_type="wounded cotyledons"
                     /dev_stage="11 day old seedlings"
                     /lab_host="DH10B"
                     /clone_lib="Gm-c1076"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     from 11 day old seedlings treated with that were treated
                     with 2 ugs/ml of a crude glucan elicitor preparation
                     isolated from the mycelial walls of Phytophthora sojae.
                     The  library was prepared using the Stratagene pBluescript
                     II SK(+) library  construction kit. Complementary DNA was
                     synthesized from mRNA using a primer  consisting of a
                     poly(dT) sequence with an XhoI restriction site. EcoRI
                     adaptors were ligated to the blunt-ended cDNA fragments
                     followed by XhoI  digestion. The cDNA fragments were
                     directionally cloned into the EcoRI-XhoI  restriction site
```

of the pBluescript vector. The ligated cDNA fragments were transformed into E.coli ElectroMax DH10B host cells. Plant material was provided by Michael G. Hahn (Complex Carbohydrate Research Center, University of Georgia) and the library was constructed by Anu Khanna (Lila Vodkin lab , University of Illinois)."
BASE COUNT 64 a 31 c 50 g 82 t
ORIGIN

Query Match 0.2%; Score 20; DB 13; Length 227;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1358 TAAATGGAAAAACAAAAAAA 1377
||||||||||||||||||||
Db 44 TAAATGGAAAAACAAAAAAA 25

RESULT 83
D79467/c
LOCUS D79467 230 bp mRNA linear EST 09-FEB-1996
DEFINITION HUM252F11B Human aorta polyA+ (TFujiwara) Homo sapiens cDNA clone GEN-252F11 5', mRNA sequence.
ACCESSION D79467
VERSION D79467.1 GI:1179818
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 230)
AUTHORS Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M., Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H., Takaichi ,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y., Maekawa,H., Shin,S. and Nakamura,Y.
TITLE Fujiwara et al. (1995)
JOURNAL Unpublished
COMMENT Contact: Tsutomu Fujiwara
Otsuka GEN Research Institute
Otsuka Pharmaceutical Co.,Ltd
463-10 kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
Tel: 0886-65-2888
Fax: 0886-37-1035.
FEATURES Location/Qualifiers
source 1. .230
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GEN-252F11"
/tissue_type="aorta"
/clone_lib="Human aorta polyA+ (TFujiwara)"
BASE COUNT 61 a 30 c 51 g 86 t 2 others
ORIGIN

Query Match 0.2%; Score 20; DB 14; Length 230;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4861 GCAAACATCAAATAATATAA 4880
||||||||||||||||||||
Db 30 GCAAACATCAAATAATATAA 11

RESULT 84
AV320570
LOCUS AV320570 233 bp mRNA linear EST 09-NOV-1999
DEFINITION AV320570 RIKEN full-length enriched, 13 days embryo male testis Mus musculus cDNA clone 6030424J06 3' similar to AF093139 Rattus norvegicus tip associating protein (TAP) mRNA, mRNA sequence.
ACCESSION AV320570
VERSION AV320570.1 GI:6290403
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 233)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai ,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata ,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al. 1999)
JOURNAL Unpublished
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y., Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki ,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .233
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="6030424J06"
/sex="male"
/tissue_type="testis"
/dev_stage="13 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 13 days embryo male testis"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
BASE COUNT 59 a 51 c 54 g 69 t
ORIGIN

```
  Query Match          0.2%;  Score 20;  DB 9;  Length 233;
  Best Local Similarity  100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6530 AGGTGCTCTGTGTACTTTAC 6549
             ||||||||||||||||||||
Db        171 AGGTGCTCTGTGTACTTTAC 190


RESULT 85
H84134/c
LOCUS       H84134                  235 bp    mRNA    linear   EST 13-NOV-1995
DEFINITION  yt02d12.s1 Soares retina N2b5HR Homo sapiens cDNA clone
            IMAGE:223127 3' similar to gb:L11672 ZINC FINGER PROTEIN 91 (HUMAN
            );; mRNA sequence.
ACCESSION   H84134
VERSION     H84134.1  GI:1062805
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 235)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 96
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 674   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 96.
FEATURES             Location/Qualifiers
     source          1. .235
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:3854161"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:223127"
                     /sex="male"
                     /tissue_type="retina"
                     /dev_stage="55 year old"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares retina N2b5HR"
                     /note="Organ: eye; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). The retinas were obtained from a 55 year old
                     Caucasian  and total cellular poly(A)+ RNA was extracted 6
                     hrs after their removal. The retina RNA  was kindly
                     provided by Roderick R. McInnes M.D. Ph.D. from the
                     University  of Toronto. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. "
BASE COUNT       72 a     42 c     38 g     77 t      6 others
ORIGIN

  Query Match          0.2%;  Score 20;  DB 14;  Length 235;
  Best Local Similarity  100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4753 TACATACTTCTCATGTTTTA 4772
              ||||||||||||||||||||
Db         21 TACATACTTCTCATGTTTTA 2


RESULT 86
AZ308464
LOCUS       AZ308464                236 bp    DNA     linear   GSS 29-SEP-2000
DEFINITION  1M0011N22F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0011N22 F, genomic survey sequence.
ACCESSION   AZ308464
VERSION     AZ308464.1  GI:10348486
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 236)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0011  row: N  column: 22
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 236.
FEATURES             Location/Qualifiers
     source          1. .236
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0011N22"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       91 a     37 c     31 g     77 t
ORIGIN

  Query Match          0.2%;  Score 20;  DB 28;  Length 236;
  Best Local Similarity  100.0%;  Pred. No. 4.3e+02;
```

```
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8610 CACTGGTTAATATACCAATA 8629
              ||||||||||||||||||||
Db        136 CACTGGTTAATATACCAATA 155


RESULT 87
AA484120/c
LOCUS       AA484120                 277 bp    mRNA    linear   EST 14-AUG-1997
DEFINITION  ne77g03.s1 NCI_CGAP_Ew1 Homo sapiens cDNA clone IMAGE:910324, mRNA
            sequence.
ACCESSION   AA484120
VERSION     AA484120.1  GI:2212933
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 277)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,
            Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 744   Std Error: 0.00
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 264.
FEATURES             Location/Qualifiers
     source          1. .277
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:910324"
                     /tissue_type="Ewing's sarcoma"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Ew1"
                     /note="Vector: pAMP10; mRNA made from Ewing's sarcoma,
                     cDNA made by oligo-dT priming.  Non-directionally cloned.
                     Size-selected on agarose gel, average insert  size 600 bp.
                     Reference:  Krizman et al. (1996) Cancer Research
                     56:5380-5383."
BASE COUNT       75 a     36 c     63 g    103 t
ORIGIN

  Query Match             0.2%;  Score 20;  DB 9;  Length 277;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
              ||||||||||||||||||||
Db         76 GCAAACATCAAATAATATAA 57


RESULT 88
AA555009/c
LOCUS       AA555009                 278 bp    mRNA    linear   EST 09-SEP-1997
DEFINITION  nl07c12.s1 NCI_CGAP_Pr11 Homo sapiens cDNA clone IMAGE:1029622,
            mRNA sequence.
ACCESSION   AA555009
VERSION     AA555009.1  GI:2325548
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 278)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Rodrigo F.
            Chuaqui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 356   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 276.
FEATURES             Location/Qualifiers
     source          1. .278
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1029622"
                     /sex="male"
                     /tissue_type="normal prostatic epithelial cells"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Pr11"
                     /note="Organ: prostate; Vector: pAMP10; mRNA made from
                     normal prostatic epithelial cells, cDNA made by oligo-dT
                     priming.  Non-directionally cloned.  Size-selected on
                     agarose gel, average insert size 600 bp.  Library made by
                     D. Krizman, NIH."
BASE COUNT       75 a     38 c     63 g    102 t
ORIGIN

  Query Match             0.2%;  Score 20;  DB 9;  Length 278;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
              ||||||||||||||||||||
Db         82 GCAAACATCAAATAATATAA 63


RESULT 89
BZ860554
LOCUS       BZ860554                 285 bp    DNA     linear   GSS 18-MAR-2003
DEFINITION  CH240_269I2.TJ CHORI-240 Bos taurus genomic clone CH240_269I2,
            genomic survey sequence.
ACCESSION   BZ860554
VERSION     BZ860554.1  GI:29087959
KEYWORDS    GSS.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 285)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Chen,D., Riggs,F., de Jong,P., Crawford,A.M.
            and McEwan,J.C.
  TITLE     Bovine BAC End Sequences from Library CHORI-240
  JOURNAL   Unpublished
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
```

```
            Clones are derived from the bovine BAC library CHORI-240
            (http://www.chori.org/bacpac/bovine240.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/ordering_information.htm). This work
            was undertaken as part of the International Bovine BAC Mapping
            Consortium (IBBMC) by AgResearch Ltd., New Zealand and The
            Institute of Genomic Research (TIGR), USA.
            Plate: 269  row: I  column: 2
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .285
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /strain="breed: Hereford"
                     /db_xref="taxon:9913"
                     /clone="CH240_269I2"
                     /sex="Male"
                     /cell_type="Blood"
                     /clone_lib="CHORI-240"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC
                     library (Male) produced by Pieter de Jong"
BASE COUNT       93 a     64 c     60 g     68 t
ORIGIN

  Query Match             0.2%;  Score 20;  DB 29;  Length 285;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6207 ACCGTGAACTTCCTGATGTC 6226
             ||||||||||||||||||||
Db        245 ACCGTGAACTTCCTGATGTC 264


RESULT 90
BE702501/c
LOCUS       BE702501                 287 bp    mRNA    linear   EST 12-SEP-2000
DEFINITION  RC1-NN1063-190700-015-g09 NN1063 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE702501
VERSION     BE702501.1  GI:10090243
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 287)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=RC1-NN1063-190
            700-015-g09&t3=2000-07-19&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 5
            High quality sequence stop: 287.
FEATURES             Location/Qualifiers
     source          1. .287
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="NN1063"
                     /note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       88 a     45 c     53 g     101 t
ORIGIN

  Query Match             0.2%;  Score 20;  DB 10;  Length 287;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
              ||||||||||||||||||||
Db        184 GCAAACATCAAATAATATAA 165


RESULT 91
AV108259/c
LOCUS       AV108259                 291 bp    mRNA    linear   EST 28-JUN-1999
DEFINITION  AV108259 Mus musculus liver C57BL/6J 13-day embryo Mus musculus
            cDNA clone 2510044B19, mRNA sequence.
ACCESSION   AV108259
VERSION     AV108259.1  GI:5255807
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
            Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara
            ,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,
            Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H., Owa,C.,
            Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara
            ,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tominaga,N.,
            Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,
            Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs
  JOURNAL   Unpublished
COMMENT     Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length cDNA
            (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="2510044B19"
                     /sex="mixed"
                     /tissue_type="liver"
```

```
                     /dev_stage="13-day embryo"
                     /clone_lib="Mus musculus liver C57BL/6J 13-day embryo"
BASE COUNT       95 a     29 c     56 g    111 t
ORIGIN

  Query Match             0.2%;  Score 20;  DB 9;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4706 AACATGTAATTAAAAAAAAT 4725
              ||||||||||||||||||||
Db         40 AACATGTAATTAAAAAAAAT 21


RESULT 92
AQ170605/c
LOCUS       AQ170605                 291 bp    DNA     linear   GSS 16-OCT-1998
DEFINITION  HS_2165_A2_F12_MF CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2165 Col=24 Row=K, genomic survey
            sequence.
ACCESSION   AQ170605
VERSION     AQ170605.1  GI:3562010
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2165  row: K  column: 24
            Class: BAC ends
            High quality sequence stop: 291.
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=2165 Col=24 Row=K"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      100 a     27 c     75 g     89 t
ORIGIN

  Query Match             0.2%;  Score 20;  DB 28;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6072 ACCTTTAAAAAATCGCAAAA 6091
              ||||||||||||||||||||
Db         27 ACCTTTAAAAAATCGCAAAA 8


RESULT 93
AA335616/c
LOCUS       AA335616                 313 bp    mRNA    linear   EST 21-APR-1997
DEFINITION  EST40039 Epididymus Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA335616
VERSION     AA335616.1  GI:1987858
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 313)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
   PUBMED   7566098
COMMENT     Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .313
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="ATCC (inhost):137285"
                     /db_xref="taxon:9606"
                     /sex="male"
                     /dev_stage="adult"
                     /clone_lib="Epididymus"
                     /note="Organ: epididymis; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
BASE COUNT       99 a     35 c     60 g    116 t      3 others
ORIGIN

  Query Match             0.2%;  Score 20;  DB 9;  Length 313;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
              ||||||||||||||||||||
Db        267 GCAAACATCAAATAATATAA 248


RESULT 94
BB275640
LOCUS       BB275640                 317 bp    mRNA    linear   EST 07-JUL-2000
DEFINITION  BB275640 RIKEN full-length enriched, 10 days neonate cortex Mus
            musculus cDNA clone A830093P14 3', mRNA sequence.
ACCESSION   BB275640
VERSION     BB275640.1  GI:8972661
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
```

```
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 317)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES              Location/Qualifiers
     source           1. .317
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /db_xref="taxon:10090"
                      /clone="A830093P14"
                      /tissue_type="cortex"
                      /dev_stage="10 days neonate"
                      /lab_host="DH10B"
                      /clone_lib="RIKEN full-length enriched, 10 days neonate
                      cortex"
                      /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                      prepared and sequenced in Mouse Genome Encyclopedia
                      Project of Genome Exploration Research Group in Riken
                      Genomic Sciences Center and Genome Science Laboratory in
                      RIKEN. Division of Experimental Animal Research in Riken
                      contributed to prepare mouse tissues. 1st strand cDNA was
                      primed with a primer [5'
                      GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                      prepared by using trehalose thermo-activated reverse
                      transcriptase and subsequently enriched for full-length by
                      cap-trapper. cDNA went through one round of normalization
                      to Rot = 20.0 and subtraction to Rot = 459.0. Second
                      strand cDNA was prepared with the primer adapter of
                      sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                      3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                      modified pBluescript KS(+) after bulk excision from Lambda
                      FLC I."
BASE COUNT      101 a     68 c     47 g    101 t
ORIGIN
```

```
  Query Match             0.2%;  Score 20;  DB 10;  Length 317;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6772 AAGAAAAAATTTTTTAACCA 6791
              ||||||||||||||||||||
Db         61 AAGAAAAAATTTTTTAACCA 80


RESULT 95
BB501942/c
LOCUS       BB501942                320 bp    mRNA    linear   EST 26-JUL-2000
DEFINITION  BB501942 RIKEN full-length enriched, 0 day neonate kidney Mus
            musculus cDNA clone D630036P08 3', mRNA sequence.
ACCESSION   BB501942
VERSION     BB501942.1  GI:9490736
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 320)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ;Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES              Location/Qualifiers
     source           1. .320
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /db_xref="taxon:10090"
                      /clone="D630036P08"
                      /tissue_type="kidney"
                      /dev_stage="0 day neonate"
                      /lab_host="DH10B"
                      /clone_lib="RIKEN full-length enriched, 0 day neonate
                      kidney"
                      /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                      prepared and sequenced in Mouse Genome Encyclopedia
```

Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 74 a 90 c 41 g 115 t
ORIGIN

Query Match 0.2%; Score 20; DB 10; Length 320;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2777 GTCATCTGAAATAATAATGG 2796
||||||||||||||||||||
Db 169 GTCATCTGAAATAATAATGG 150

RESULT 96
AA665355/c
LOCUS AA665355 347 bp mRNA linear EST 18-FEB-1998
DEFINITION nt89e11.s1 NCI_CGAP_Pr12 Homo sapiens cDNA clone IMAGE:1205708, mRNA sequence.
ACCESSION AA665355
VERSION AA665355.1 GI:2619968
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 347)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D., Rodrigo F. Chuaqui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 459 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 342.
FEATURES Location/Qualifiers
source 1. .347
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1205708"
/sex="male"
/tissue_type="metastatic prostate bone lesion"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Pr12"
/note="Vector: pAMP10; mRNA made from metastatic prostate lesion of the bone, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Library made by D. Krizman, NIH."
BASE COUNT 103 a 51 c 69 g 124 t
ORIGIN

Query Match 0.2%; Score 20; DB 9; Length 347;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4861 GCAAACATCAAATAATATAA 4880
||||||||||||||||||||
Db 57 GCAAACATCAAATAATATAA 38

RESULT 97
BF477423
LOCUS BF477423 366 bp mRNA linear EST 05-DEC-2000
DEFINITION nac60e12.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3438815 3', mRNA sequence.
ACCESSION BF477423
VERSION BF477423.1 GI:11548250
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 366)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGAP), Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 366.
FEATURES Location/Qualifiers
source 1. .366
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3438815"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Brn23"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCATATCTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 135 a 68 c 51 g 112 t
ORIGIN

Query Match 0.2%; Score 20; DB 10; Length 366;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4861 GCAAACATCAAATAATATAA 4880
||||||||||||||||||||
Db 338 GCAAACATCAAATAATATAA 357

RESULT 98
AI024142
LOCUS AI024142 371 bp mRNA linear EST 18-JUN-1998

```
DEFINITION  ov73a07.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1642932
            3', mRNA sequence.
ACCESSION   AI024142
VERSION     AI024142.1  GI:3239186
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 371)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 318.
FEATURES             Location/Qualifiers
     source          1. .371
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1642932"
                     /sex="male"
                     /lab_host="DH10B"
                     /clone_lib="Soares_testis_NHT"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. 
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      130 a     66 c     53 g    122 t
ORIGIN

  Query Match             0.2%;  Score 20;  DB 9;  Length 371;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
              ||||||||||||||||||||
Db        351 GCAAACATCAAATAATATAA 370


RESULT 99
AA973137/c
LOCUS       AA973137                373 bp    mRNA    linear   EST 20-MAY-1998
DEFINITION  on93b01.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1564201 3' similar to TR:P97657 P97657 ZINC FINGER PROTEIN 1
            ;, mRNA sequence.
ACCESSION   AA973137
VERSION     AA973137.1  GI:3148317
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 373)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 319.
FEATURES             Location/Qualifiers
     source          1. .373
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1564201"
                     /lab_host="DH10B"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT      113 a     58 c     59 g    143 t
ORIGIN

  Query Match             0.2%;  Score 20;  DB 9;  Length 373;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4753 TACATACTTCTCATGTTTTA 4772
              ||||||||||||||||||||
Db         44 TACATACTTCTCATGTTTTA 25


RESULT 100
AA664116
LOCUS       AA664116                374 bp    mRNA    linear   EST 12-NOV-1997
DEFINITION  ac06b06.s1 Stratagene lung (#937210) Homo sapiens cDNA clone
            IMAGE:855635 3', mRNA sequence.
ACCESSION   AA664116
VERSION     AA664116.1  GI:2618107
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 374)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .374
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:855635"
```

/sex="male"
/dev_stage="72 years"
/lab_host="SOLR cells (kanamycin resistant)"
/clone_lib="Stratagene lung (#937210)"
/note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dT. normal lung. Average insert size: 1.0 kb; Uni-ZAP XR
Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3'
adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT 132 a 68 c 52 g 122 t
ORIGIN

Query Match 0.2%; Score 20; DB 9; Length 374;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4861 GCAAACATCAAATAATATAA 4880
||||||||||||||||||||
Db 351 GCAAACATCAAATAATATAA 370

Search completed: August 22, 2003, 14:38:54
Job time : 16008 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 22, 2003, 00:42:37 ; Search time 533 Seconds
(without alignments)
7513.451 Million cell updates/sec

Title: US-09-831-000-1_COPY_60966_70038
Perfect score: 9073
Sequence: 1 atggagagttccgtcggatg.....gcggccacggatcgccttga 9073

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 300 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	21	0.2	124884	4	US-09-661-596A-76	Sequence 76, Appl
	2	20	0.2	249	4	US-09-252-991A-9860	Sequence 9860, Ap
c	3	20	0.2	960	4	US-09-339-159B-21	Sequence 21, Appl
c	4	20	0.2	2017	4	US-09-690-454-45	Sequence 45, Appl
c	5	20	0.2	2026	4	US-09-690-454-30	Sequence 30, Appl
	6	20	0.2	2871	4	US-09-107-532A-1353	Sequence 1353, Ap
	7	20	0.2	35100	2	US-08-770-379-18	Sequence 18, Appl
	8	20	0.2	35100	3	US-08-757-669A-18	Sequence 18, Appl
	9	20	0.2	35100	4	US-09-230-371A-18	Sequence 18, Appl
	10	20	0.2	1230025	4	US-09-198-452A-1	Sequence 1, Appli
c	11	19	0.2	234	4	US-09-328-352-2598	Sequence 2598, Ap
	12	19	0.2	284	4	US-09-313-294A-4581	Sequence 4581, Ap
	13	19	0.2	292	4	US-09-313-294A-88	Sequence 88, Appl
	14	19	0.2	471	4	US-09-328-352-2603	Sequence 2603, Ap
c	15	19	0.2	719	4	US-09-620-312D-669	Sequence 669, App
	16	19	0.2	989	1	US-07-885-970A-7	Sequence 7, Appli
	17	19	0.2	989	1	US-08-298-687A-7	Sequence 7, Appli
	18	19	0.2	989	1	US-08-530-797-6	Sequence 6, Appli
	19	19	0.2	989	1	US-08-298-829-7	Sequence 7, Appli
	20	19	0.2	989	2	US-08-787-335-6	Sequence 6, Appli
	21	19	0.2	1215	4	US-09-328-352-49	Sequence 49, Appl
c	22	19	0.2	1412	1	US-08-299-849B-21	Sequence 21, Appl
c	23	19	0.2	1412	2	US-08-142-368A-21	Sequence 21, Appl
c	24	19	0.2	1412	3	US-08-967-727-21	Sequence 21, Appl
c	25	19	0.2	1412	3	US-08-037-230D-21	Sequence 21, Appl
c	26	19	0.2	1412	4	US-09-583-850-21	Sequence 21, Appl
c	27	19	0.2	1412	4	US-09-579-197-21	Sequence 21, Appl
c	28	19	0.2	1412	4	US-09-404-026-21	Sequence 21, Appl
c	29	19	0.2	1511	4	US-08-979-847B-117	Sequence 117, App
	30	19	0.2	2167	2	US-08-461-775-9	Sequence 9, Appli
	31	19	0.2	2167	3	US-09-031-606-9	Sequence 9, Appli
	32	19	0.2	2668	2	US-08-461-775-11	Sequence 11, Appl
	33	19	0.2	2668	3	US-09-031-606-11	Sequence 11, Appl
c	34	19	0.2	2931	3	US-09-056-105-15	Sequence 15, Appl
	35	19	0.2	4386	4	US-09-300-008B-1	Sequence 1, Appli
c	36	18	0.2	36	4	US-09-177-419C-11	Sequence 11, Appl
c	37	18	0.2	645	4	US-09-134-001C-1986	Sequence 1986, Ap
	38	18	0.2	696	4	US-09-134-001C-987	Sequence 987, App
	39	18	0.2	792	1	US-08-203-905B-3	Sequence 3, Appli
	40	18	0.2	894	2	US-08-467-963C-28	Sequence 28, Appl
	41	18	0.2	894	2	US-08-838-189D-28	Sequence 28, Appl
	42	18	0.2	894	3	US-08-852-344D-28	Sequence 28, Appl
	43	18	0.2	894	3	US-08-344-639E-28	Sequence 28, Appl
	44	18	0.2	920	2	US-08-467-963C-7	Sequence 7, Appli
	45	18	0.2	920	2	US-08-838-189D-7	Sequence 7, Appli
	46	18	0.2	920	3	US-08-852-344D-7	Sequence 7, Appli
	47	18	0.2	920	3	US-08-344-639E-7	Sequence 7, Appli
	48	18	0.2	920	3	US-08-467-969A-7	Sequence 7, Appli
	49	18	0.2	920	3	US-08-467-961A-7	Sequence 7, Appli
	50	18	0.2	920	3	US-08-001-554A-7	Sequence 7, Appli
c	51	18	0.2	997	3	US-09-057-860A-3	Sequence 3, Appli
c	52	18	0.2	1239	4	US-09-252-991A-1942	Sequence 1942, Ap
	53	18	0.2	1336	4	US-09-149-476-230	Sequence 230, App
	54	18	0.2	1442	4	US-08-936-165A-67	Sequence 67, Appl
c	55	18	0.2	1541	4	US-09-149-476-65	Sequence 65, Appl
c	56	18	0.2	1781	4	US-09-818-512-1	Sequence 1, Appli
	57	18	0.2	2245	1	US-08-203-905B-1	Sequence 1, Appli
	58	18	0.2	2525	4	US-09-482-273-20	Sequence 20, Appl
c	59	18	0.2	2598	4	US-09-252-991A-2112	Sequence 2112, Ap
	60	18	0.2	2844	3	US-08-815-809-2	Sequence 2, Appli
	61	18	0.2	2856	2	US-08-816-155B-4	Sequence 4, Appli
	62	18	0.2	2856	3	US-09-079-587-4	Sequence 4, Appli
c	63	18	0.2	2866	4	US-09-724-864-30	Sequence 30, Appl
c	64	18	0.2	2951	4	US-09-177-419C-3	Sequence 3, Appli
	65	18	0.2	3291	1	US-08-021-601-1	Sequence 1, Appli
	66	18	0.2	3291	1	US-08-082-849B-1	Sequence 1, Appli
	67	18	0.2	3291	5	PCT-US94-01624-1	Sequence 1, Appli
	68	18	0.2	3636	4	US-09-134-001C-1983	Sequence 1983, Ap
c	69	18	0.2	3963	1	US-08-464-961-1	Sequence 1, Appli
c	70	18	0.2	3963	3	US-08-907-800A-1	Sequence 1, Appli
c	71	18	0.2	3963	4	US-09-083-541-1	Sequence 1, Appli
c	72	18	0.2	3963	4	US-08-969-315-1	Sequence 1, Appli
c	73	18	0.2	3963	5	PCT-US96-08233-1	Sequence 1, Appli
	74	18	0.2	4032	1	US-08-107-748-3	Sequence 3, Appli
	75	18	0.2	4032	1	US-08-245-809-4	Sequence 4, Appli
	76	18	0.2	4032	5	PCT-US92-01385-3	Sequence 3, Appli
c	77	18	0.2	4752	1	US-08-201-697-3	Sequence 3, Appli
c	78	18	0.2	4799	1	US-08-201-697-6	Sequence 6, Appli
c	79	18	0.2	4815	1	US-08-201-697-5	Sequence 5, Appli
c	80	18	0.2	5679	1	US-08-201-697-1	Sequence 1, Appli
c	81	18	0.2	6628	3	US-08-815-809-3	Sequence 3, Appli
c	82	18	0.2	6649	2	US-08-816-155B-5	Sequence 5, Appli
c	83	18	0.2	6649	3	US-09-079-587-5	Sequence 5, Appli
	84	18	0.2	7970	1	US-08-135-511-31	Sequence 31, Appl
	85	18	0.2	7997	1	US-08-187-453-31	Sequence 31, Appl
	86	18	0.2	7997	1	US-08-562-985A-7	Sequence 7, Appli
	87	18	0.2	10726	4	US-08-961-527-66	Sequence 66, Appl
	88	18	0.2	13206	4	US-08-961-527-33	Sequence 33, Appl
	89	18	0.2	44377	2	US-08-804-227C-7	Sequence 7, Appli
	90	18	0.2	44377	2	US-08-804-198-1	Sequence 1, Appli
	91	18	0.2	49377	1	US-08-764-233A-1	Sequence 1, Appli
c	92	18	0.2	90050	3	US-09-245-041-5	Sequence 5, Appli
c	93	18	0.2	116592	4	US-09-818-512-3	Sequence 3, Appli
	94	18	0.2	152331	3	US-09-128-155-16	Sequence 16, Appl
	95	18	0.2	580073	4	US-08-545-528D-1	Sequence 1, Appli
	96	18	0.2	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c	97	18	0.2	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	98	18	0.2	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	99	17	0.2	38	1	US-08-173-510B-57	Sequence 57, Appl
c	100	17	0.2	38	1	US-08-458-218-57	Sequence 57, Appl

c 101	17	0.2	38	2	US-08-450-497-57	Sequence 57, Appl
102	17	0.2	155	4	US-09-051-239A-7	Sequence 7, Appli
c 103	17	0.2	251	2	US-08-623-906A-16	Sequence 16, Appl
104	17	0.2	266	4	US-09-313-294A-576	Sequence 576, App
105	17	0.2	270	4	US-09-107-532A-2972	Sequence 2972, Ap
c 106	17	0.2	294	4	US-09-313-294A-7136	Sequence 7136, Ap
c 107	17	0.2	300	4	US-09-205-258-128	Sequence 128, App
108	17	0.2	371	4	US-09-643-597-150	Sequence 150, App
109	17	0.2	371	4	US-09-480-884A-150	Sequence 150, App
110	17	0.2	371	4	US-09-542-615A-150	Sequence 150, App
111	17	0.2	371	4	US-09-606-421B-150	Sequence 150, App
112	17	0.2	375	4	US-09-134-001C-93	Sequence 93, Appl
c 113	17	0.2	375	4	US-09-252-991A-9279	Sequence 9279, Ap
114	17	0.2	407	4	US-09-439-313-402	Sequence 402, App
115	17	0.2	407	4	US-09-352-616A-402	Sequence 402, App
c 116	17	0.2	453	4	US-09-134-001C-2323	Sequence 2323, Ap
c 117	17	0.2	477	3	US-08-706-391B-6	Sequence 6, Appli
118	17	0.2	495	4	US-09-252-991A-159	Sequence 159, App
119	17	0.2	528	3	US-09-157-177-111	Sequence 111, App
c 120	17	0.2	534	4	US-09-252-991A-10792	Sequence 10792, A
c 121	17	0.2	540	4	US-09-702-705-1370	Sequence 1370, Ap
c 122	17	0.2	540	4	US-09-736-457-1370	Sequence 1370, Ap
123	17	0.2	543	4	US-09-252-991A-673	Sequence 673, App
124	17	0.2	551	3	US-08-896-164-2	Sequence 2, Appli
125	17	0.2	582	4	US-09-252-991A-624	Sequence 624, App
126	17	0.2	597	4	US-09-252-991A-11140	Sequence 11140, A
127	17	0.2	600	3	US-08-896-164-8	Sequence 8, Appli
128	17	0.2	623	3	US-09-385-982-167	Sequence 167, App
c 129	17	0.2	639	4	US-09-252-991A-9950	Sequence 9950, Ap
130	17	0.2	639	4	US-09-328-352-1526	Sequence 1526, Ap
c 131	17	0.2	642	4	US-09-252-991A-709	Sequence 709, App
132	17	0.2	648	4	US-09-702-705-609	Sequence 609, App
133	17	0.2	648	4	US-09-736-457-609	Sequence 609, App
c 134	17	0.2	684	4	US-09-149-476-82	Sequence 82, Appl
135	17	0.2	688	4	US-09-702-705-186	Sequence 186, App
136	17	0.2	688	4	US-09-736-457-186	Sequence 186, App
137	17	0.2	693	1	US-08-168-091A-3	Sequence 3, Appli
138	17	0.2	707	3	US-08-896-164-4	Sequence 4, Appli
139	17	0.2	744	4	US-09-671-317-392	Sequence 392, App
c 140	17	0.2	771	4	US-09-252-991A-771	Sequence 771, App
141	17	0.2	789	4	US-09-252-991A-5788	Sequence 5788, Ap
142	17	0.2	807	4	US-09-252-991A-164	Sequence 164, App
143	17	0.2	813	4	US-09-252-991A-9300	Sequence 9300, Ap
144	17	0.2	835	3	US-08-896-164-9	Sequence 9, Appli
145	17	0.2	849	4	US-09-107-532A-2666	Sequence 2666, Ap
146	17	0.2	852	4	US-09-599-360B-9	Sequence 9, Appli
147	17	0.2	852	4	US-09-599-360B-61	Sequence 61, Appl
148	17	0.2	862	2	US-08-222-719-17	Sequence 17, Appl
149	17	0.2	862	2	US-08-470-925-17	Sequence 17, Appl
150	17	0.2	862	2	US-08-471-613-17	Sequence 17, Appl
151	17	0.2	862	5	PCT-US93-10443-17	Sequence 17, Appl
c 152	17	0.2	873	4	US-09-252-991A-10913	Sequence 10913, A
c 153	17	0.2	876	4	US-09-328-352-1152	Sequence 1152, Ap
c 154	17	0.2	877	1	US-08-225-473-1	Sequence 1, Appli
c 155	17	0.2	891	4	US-09-252-991A-176	Sequence 176, App
156	17	0.2	917	4	US-09-016-434-681	Sequence 681, App
c 157	17	0.2	951	4	US-09-252-991A-7708	Sequence 7708, Ap
158	17	0.2	957	4	US-09-252-991A-5803	Sequence 5803, Ap
159	17	0.2	975	3	US-09-216-001-2	Sequence 2, Appli
160	17	0.2	975	3	US-08-878-862-2	Sequence 2, Appli
161	17	0.2	981	4	US-09-252-991A-7890	Sequence 7890, Ap
c 162	17	0.2	990	4	US-09-641-638-277	Sequence 277, App
c 163	17	0.2	999	4	US-09-252-991A-182	Sequence 182, App
164	17	0.2	1001	4	US-09-671-317-393	Sequence 393, App
c 165	17	0.2	1002	4	US-09-641-638-587	Sequence 587, App
166	17	0.2	1008	4	US-09-252-991A-7923	Sequence 7923, Ap
c 167	17	0.2	1024	4	US-09-328-475C-24	Sequence 24, Appl
168	17	0.2	1086	1	US-08-255-670A-3	Sequence 3, Appli
c 169	17	0.2	1122	4	US-09-328-352-640	Sequence 640, App
170	17	0.2	1155	4	US-09-328-352-1775	Sequence 1775, Ap
171	17	0.2	1176	4	US-09-252-991A-154	Sequence 154, App
172	17	0.2	1206	4	US-09-107-532A-224	Sequence 224, App
173	17	0.2	1227	4	US-09-252-991A-7853	Sequence 7853, Ap
c 174	17	0.2	1239	4	US-09-328-352-346	Sequence 346, App
175	17	0.2	1248	3	US-09-105-537-7	Sequence 7, Appli
176	17	0.2	1248	4	US-09-252-991A-5821	Sequence 5821, Ap
c 177	17	0.2	1282	4	US-09-221-017B-310	Sequence 310, App
178	17	0.2	1314	4	US-09-252-991A-10130	Sequence 10130, A
c 179	17	0.2	1332	4	US-09-252-991A-10974	Sequence 10974, A
180	17	0.2	1409	2	US-08-598-873-69	Sequence 69, Appl
181	17	0.2	1409	3	US-08-605-430-69	Sequence 69, Appl
182	17	0.2	1413	4	US-09-252-991A-11269	Sequence 11269, A
c 183	17	0.2	1497	4	US-09-252-991A-10612	Sequence 10612, A
c 184	17	0.2	1500	1	US-08-704-398-1	Sequence 1, Appli
c 185	17	0.2	1500	5	PCT-US95-05966-1	Sequence 1, Appli
c 186	17	0.2	1542	4	US-09-252-991A-7625	Sequence 7625, Ap
187	17	0.2	1624	1	US-08-255-670A-1	Sequence 1, Appli
c 188	17	0.2	1653	4	US-09-252-991A-7554	Sequence 7554, Ap
189	17	0.2	1662	4	US-09-504-358-5	Sequence 5, Appli
190	17	0.2	1662	4	US-09-954-314-5	Sequence 5, Appli
191	17	0.2	1701	4	US-09-252-991A-7927	Sequence 7927, Ap
c 192	17	0.2	1835	4	US-09-417-704-2	Sequence 2, Appli
193	17	0.2	1929	3	US-08-750-180-1	Sequence 1, Appli
c 194	17	0.2	1992	4	US-09-484-970B-150	Sequence 150, App
c 195	17	0.2	2220	4	US-09-328-352-1922	Sequence 1922, Ap
196	17	0.2	2266	2	US-08-724-394A-18	Sequence 18, Appl
197	17	0.2	2511	4	US-09-620-312D-559	Sequence 559, App
198	17	0.2	2580	4	US-09-620-312D-813	Sequence 813, App
199	17	0.2	2677	4	US-09-221-017B-959	Sequence 959, App
c 200	17	0.2	2739	4	US-09-252-991A-7592	Sequence 7592, Ap
201	17	0.2	2778	4	US-09-252-991A-7958	Sequence 7958, Ap
c 202	17	0.2	2850	4	US-09-184-964-2	Sequence 2, Appli
c 203	17	0.2	2938	2	US-08-343-443B-3	Sequence 3, Appli
c 204	17	0.2	2959	1	US-08-408-188A-1	Sequence 1, Appli
c 205	17	0.2	2959	1	US-08-408-188A-2	Sequence 2, Appli
c 206	17	0.2	2959	1	US-08-408-188A-3	Sequence 3, Appli
c 207	17	0.2	2959	1	US-08-408-188A-4	Sequence 4, Appli
c 208	17	0.2	2959	1	US-08-766-488-1	Sequence 1, Appli
c 209	17	0.2	2959	1	US-08-766-488-2	Sequence 2, Appli
c 210	17	0.2	2959	1	US-08-766-488-3	Sequence 3, Appli
c 211	17	0.2	2959	1	US-08-766-488-4	Sequence 4, Appli
c 212	17	0.2	2995	1	US-08-592-126-85	Sequence 85, Appl
c 213	17	0.2	2995	4	US-09-168-595-85	Sequence 85, Appl
214	17	0.2	3061	2	US-08-692-787-47	Sequence 47, Appl
215	17	0.2	3061	3	US-09-097-199-47	Sequence 47, Appl
c 216	17	0.2	3181	4	US-09-620-312D-856	Sequence 856, App
217	17	0.2	3317	4	US-09-193-562D-1	Sequence 1, Appli
218	17	0.2	3418	4	US-09-193-562D-29	Sequence 29, Appl
219	17	0.2	3639	2	US-08-737-524B-26	Sequence 26, Appl
220	17	0.2	3806	4	US-09-620-312D-608	Sequence 608, App
221	17	0.2	4285	4	US-09-410-464-1	Sequence 1, Appli
c 222	17	0.2	4429	1	US-08-308-872B-3	Sequence 3, Appli
c 223	17	0.2	4435	1	US-08-308-872B-5	Sequence 5, Appli
224	17	0.2	4582	2	US-08-993-228-9	Sequence 9, Appli
225	17	0.2	4665	4	US-09-252-991A-10243	Sequence 10243, A
226	17	0.2	5437	1	US-07-661-610C-1	Sequence 1, Appli
227	17	0.2	5746	2	US-08-637-759B-38	Sequence 38, Appl
228	17	0.2	5746	3	US-08-871-355A-38	Sequence 38, Appl
229	17	0.2	5746	4	US-09-201-945-38	Sequence 38, Appl
c 230	17	0.2	5970	3	US-09-320-878-21	Sequence 21, Appl
c 231	17	0.2	5970	4	US-09-141-908-11	Sequence 11, Appl
c 232	17	0.2	5970	4	US-09-657-440-21	Sequence 21, Appl
233	17	0.2	6572	4	US-09-620-312D-823	Sequence 823, App
234	17	0.2	7605	3	US-09-417-455-8	Sequence 8, Appli
235	17	0.2	7605	4	US-09-348-942-8	Sequence 8, Appli
236	17	0.2	7605	4	US-09-457-626-8	Sequence 8, Appli
237	17	0.2	7605	4	US-09-576-008-8	Sequence 8, Appli
c 238	17	0.2	7785	2	US-08-276-967-1	Sequence 1, Appli
239	17	0.2	7902	4	US-08-961-527-112	Sequence 112, App
c 240	17	0.2	8512	4	US-08-961-527-129	Sequence 129, App
c 241	17	0.2	9541	4	US-08-961-527-132	Sequence 132, App
242	17	0.2	9873	4	US-09-328-352-1360	Sequence 1360, Ap
243	17	0.2	10380	3	US-09-077-354B-3	Sequence 3, Appli
244	17	0.2	10629	4	US-09-504-358-15	Sequence 15, Appl
245	17	0.2	10629	4	US-09-954-314-15	Sequence 15, Appl
246	17	0.2	11091	4	US-09-134-001C-2243	Sequence 2243, Ap

c 247	17	0.2	11770	4	US-08-961-527-172	Sequence 172, App
c 248	17	0.2	13613	3	US-09-105-537-3	Sequence 3, Appli
c 249	17	0.2	13875	2	US-08-734-344-1	Sequence 1, Appli
250	17	0.2	14273	4	US-08-961-527-40	Sequence 40, Appl
c 251	17	0.2	20137	3	US-09-262-773-206	Sequence 206, App
c 252	17	0.2	20138	3	US-09-262-773-9	Sequence 9, Appli
c 253	17	0.2	23071	3	US-09-262-773-210	Sequence 210, App
254	17	0.2	28958	1	US-08-258-261B-6	Sequence 6, Appli
255	17	0.2	28958	1	US-08-456-837-6	Sequence 6, Appli
256	17	0.2	28958	1	US-08-457-342-6	Sequence 6, Appli
257	17	0.2	28958	1	US-08-457-646A-6	Sequence 6, Appli
258	17	0.2	28958	1	US-08-458-076A-6	Sequence 6, Appli
259	17	0.2	28958	1	US-08-764-233A-4	Sequence 4, Appli
260	17	0.2	28958	1	US-08-457-335A-6	Sequence 6, Appli
261	17	0.2	28958	1	US-08-729-214-6	Sequence 6, Appli
262	17	0.2	28958	3	US-09-028-934-6	Sequence 6, Appli
263	17	0.2	31208	4	US-09-852-067-3	Sequence 3, Appli
264	17	0.2	35100	2	US-08-770-379-19	Sequence 19, Appl
265	17	0.2	35100	3	US-08-757-669A-19	Sequence 19, Appl
266	17	0.2	35100	4	US-09-230-371A-19	Sequence 19, Appl
c 267	17	0.2	36800	4	US-08-311-731A-139	Sequence 139, App
268	17	0.2	40000	4	US-09-780-049-18	Sequence 18, Appl
c 269	17	0.2	40328	3	US-08-742-185-102	Sequence 102, App
270	17	0.2	43804	4	US-09-171-461-1	Sequence 1, Appli
c 271	17	0.2	53332	4	US-09-801-861-3	Sequence 3, Appli
272	17	0.2	66804	4	US-09-740-041-3	Sequence 3, Appli
c 273	17	0.2	162450	4	US-09-345-882-1	Sequence 1, Appli
c 274	17	0.2	168575	4	US-09-426-290-1	Sequence 1, Appli
275	17	0.2	202001	4	US-09-734-674-3	Sequence 3, Appli
c 276	17	0.2	246240	2	US-08-724-394A-20	Sequence 20, Appl
c 277	17	0.2	246240	2	US-08-724-394A-21	Sequence 21, Appl
c 278	17	0.2	246240	2	US-08-724-394A-22	Sequence 22, Appl
c 279	17	0.2	319608	4	US-09-539-333D-1	Sequence 1, Appli
c 280	17	0.2	319608	4	US-09-679-409-1	Sequence 1, Appli
281	17	0.2	536165	4	US-09-214-808-1	Sequence 1, Appli
c 282	17	0.2	580073	4	US-08-545-528D-1	Sequence 1, Appli
c 283	17	0.2	1230025	4	US-09-198-452A-1	Sequence 1, Appli
c 284	17	0.2	1664976	4	US-08-916-421B-1	Sequence 1, Appli
285	17	0.2	1830121	4	US-09-557-884-1	Sequence 1, Appli
286	17	0.2	1830121	4	US-09-643-990A-1	Sequence 1, Appli
287	16	0.2	33	1	US-08-094-534-20	Sequence 20, Appl
288	16	0.2	33	2	US-08-581-543-20	Sequence 20, Appl
289	16	0.2	33	5	PCT-US94-08000-20	Sequence 20, Appl
290	16	0.2	36	2	US-08-535-494-11	Sequence 11, Appl
291	16	0.2	36	3	US-09-097-635-11	Sequence 11, Appl
c 292	16	0.2	48	1	US-08-093-453B-7	Sequence 7, Appli
c 293	16	0.2	48	1	US-08-459-041A-4	Sequence 4, Appli
c 294	16	0.2	50	3	US-09-042-353-9	Sequence 9, Appli
c 295	16	0.2	50	4	US-08-758-417A-277	Sequence 277, App
296	16	0.2	55	4	US-09-180-422B-20	Sequence 20, Appl
c 297	16	0.2	55	4	US-09-180-422B-21	Sequence 21, Appl
c 298	16	0.2	62	1	US-08-443-568B-29	Sequence 29, Appl
c 299	16	0.2	62	5	PCT-US94-06997-29	Sequence 29, Appl
c 300	16	0.2	63	1	US-08-175-155-61	Sequence 61, Appl

ALIGNMENTS

```
RESULT 1
US-09-661-596A-76
; Sequence 76, Application US/09661596A
; Patent No. 6528066
; GENERAL INFORMATION:
;  APPLICANT: Grose, Charles
;  APPLICANT:  Santos, Richard
;  TITLE OF INVENTION: VARIANT VARICELLA-ZOSTER VIRUSES AND METHODS OF USE
;  FILE REFERENCE: 140.0011 0101
;  CURRENT APPLICATION NUMBER: US/09/661,596A
;  CURRENT FILING DATE:  2000-09-14
;  PRIOR APPLICATION NUMBER: US 60/153,779
;  PRIOR FILING DATE: 1999-09-14
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 76
;   LENGTH: 124884
;   TYPE: DNA
;   ORGANISM: Varicella zoster
US-09-661-596A-76

  Query Match             0.2%;  Score 21;  DB 4;  Length 124884;
  Best Local Similarity   100.0%;  Pred. No. 3.6;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        711 CATAGTATGCGTCGGTTCGCC 731
              |||||||||||||||||||||
Db      96832 CATAGTATGCGTCGGTTCGCC 96852


RESULT 2
US-09-252-991A-9860
; Sequence 9860, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 9860
;   LENGTH: 249
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9860

  Query Match             0.2%;  Score 20;  DB 4;  Length 249;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1641 CCTGGAAGACGACATCTTCT 1660
              ||||||||||||||||||||
Db        201 CCTGGAAGACGACATCTTCT 220


RESULT 3
US-09-339-159B-21/c
; Sequence 21, Application US/09339159B
; Patent No. 6566114
; GENERAL INFORMATION:
;  APPLICANT: Kauppinen, Markus
;  APPLICANT:  Schulein, Martin
;  APPLICANT:  Schnorr, Kirk
;  APPLICANT:  Andersen, Lene
;  APPLICANT:  Bjornvad, Mads
;  TITLE OF INVENTION: No. 6566114el Mannanases
;  FILE REFERENCE: 5440.204-US
;  CURRENT APPLICATION NUMBER: US/09/339,159B
;  CURRENT FILING DATE:  1999-06-24
;  NUMBER OF SEQ ID NOS: 55
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
;   LENGTH: 960
;   TYPE: DNA
;   ORGANISM: Bacillus sp.
US-09-339-159B-21

  Query Match             0.2%;  Score 20;  DB 4;  Length 960;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       2392 ATGGAAAACATTGTGTTTTT 2411
              ||||||||||||||||||||
Db        653 ATGGAAAACATTGTGTTTTT 634


RESULT 4
US-09-690-454-45/c
; Sequence 45, Application US/09690454
; Patent No. 6531447
; GENERAL INFORMATION:
;  APPLICANT: Steven M. Ruben, et al.
;  TITLE OF INVENTION: 32 Human Secreted Proteins
;  FILE REFERENCE: PZ006P1
;  CURRENT APPLICATION NUMBER: US/09/690,454
;  CURRENT FILING DATE:  2000-10-18
;  PRIOR APPLICATION NUMBER: 09/189,144
;  PRIOR FILING DATE: 1998-11-10
;  PRIOR APPLICATION NUMBER: 60/044,039
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/048,093
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/048,190
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/050,935
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/048,101
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/048,356
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/056,250
;  PRIOR FILING DATE: August 29, 1997
;  PRIOR APPLICATION NUMBER: 60/056,296
;  PRIOR FILING DATE: August 29, 1997
;  PRIOR APPLICATION NUMBER: 60/056,293
;  PRIOR FILING DATE: August 29, 1997
;  NUMBER OF SEQ ID NOS: 229
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
;   LENGTH: 2017
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-690-454-45

  Query Match             0.2%;  Score 20;  DB 4;  Length 2017;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
              ||||||||||||||||||||
Db       1612 GCAAACATCAAATAATATAA 1593


RESULT 5
US-09-690-454-30/c
; Sequence 30, Application US/09690454
; Patent No. 6531447
; GENERAL INFORMATION:
;  APPLICANT: Steven M. Ruben, et al.
;  TITLE OF INVENTION: 32 Human Secreted Proteins
;  FILE REFERENCE: PZ006P1
;  CURRENT APPLICATION NUMBER: US/09/690,454
;  CURRENT FILING DATE:  2000-10-18
;  PRIOR APPLICATION NUMBER: 09/189,144
;  PRIOR FILING DATE: 1998-11-10
;  PRIOR APPLICATION NUMBER: 60/044,039
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/048,093
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/048,190
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/050,935
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/048,101
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/048,356
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/056,250
;  PRIOR FILING DATE: August 29, 1997
;  PRIOR APPLICATION NUMBER: 60/056,296
;  PRIOR FILING DATE: August 29, 1997
;  PRIOR APPLICATION NUMBER: 60/056,293
;  PRIOR FILING DATE: August 29, 1997
;  NUMBER OF SEQ ID NOS: 229
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
;   LENGTH: 2026
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-690-454-30

  Query Match             0.2%;  Score 20;  DB 4;  Length 2026;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
              ||||||||||||||||||||
Db       1620 GCAAACATCAAATAATATAA 1601


RESULT 6
US-09-107-532A-1353
; Sequence 1353, Application US/09107532A
; Patent No. 6583275
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                            ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;        NUMBER OF SEQUENCES: 7310
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: PC
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: ASCII
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,532A
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/085,598
;             FILING DATE: 14 May 1998
;             APPLICATION NUMBER: 60/051571
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-012
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 1353:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 2871 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;        MOLECULE TYPE: DNA (genomic)
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        ORIGINAL SOURCE:
;             ORGANISM: Enterococcus faecium
```

```
;       FEATURE:
;            NAME/KEY:  misc_feature
;            LOCATION: (B) LOCATION 1...2871
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1353:
US-09-107-532A-1353

  Query Match             0.2%;  Score 20;  DB 4;  Length 2871;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6710 CAAAGATTTAAAAACACGTT 6729
             ||||||||||||||||||||
Db       2355 CAAAGATTTAAAAACACGTT 2374


RESULT 7
US-08-770-379-18
; Sequence 18, Application US/08770379
; Patent No. 5849564
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
;    TITLE OF INVENTION:  HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,379
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  52342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  35100 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-770-379-18

  Query Match             0.2%;  Score 20;  DB 2;  Length 35100;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3776 GTGAACCAATCCCAGCCAAG 3795
             ||||||||||||||||||||
Db      33802 GTGAACCAATCCCAGCCAAG 33821


RESULT 8
US-08-757-669A-18
; Sequence 18, Application US/08757669A
; Patent No. 6183751
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
;    TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/757,669A
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  45185-F
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  35100 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-757-669A-18

  Query Match             0.2%;  Score 20;  DB 3;  Length 35100;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3776 GTGAACCAATCCCAGCCAAG 3795
             ||||||||||||||||||||
Db      33802 GTGAACCAATCCCAGCCAAG 33821


RESULT 9
US-09-230-371A-18
; Sequence 18, Application US/09230371A
; Patent No. 6348586
; GENERAL INFORMATION:
;  APPLICANT:  Chang, Yuan
;  APPLICANT:  Bohenzky, Roy A
;  APPLICANT:  Russo, James J
;  APPLICANT:  Edelman, Isidore S
;  APPLICANT:  Moore, Patrick S
;  TITLE OF INVENTION:  UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE:  45185-G-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/230,371A
;  CURRENT FILING DATE:  1999-11-17
;  PRIOR APPLICATION NUMBER:  PCT/US97/13346
;  PRIOR FILING DATE:  1997-07-22
;  NUMBER OF SEQ ID NOS:  30
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 18
;   LENGTH: 35100
;   TYPE: DNA
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
```

US-09-230-371A-18

Query Match 0.2%; Score 20; DB 4; Length 35100;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       3776 GTGAACCAATCCCAGCCAAG 3795
              ||||||||||||||||||||
Db      33802 GTGAACCAATCCCAGCCAAG 33821
```

RESULT 10
US-09-198-452A-1
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, preve
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
; LENGTH: 1230025
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(15000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (15001)..(30000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (30001)..(45000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (45001)..(60000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (60001)..(75000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (75001)..(90000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (90001)..(105000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (105001)..(120000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (120001)..(135000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (135001)..(150000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (150001)..(165000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (165001)..(180000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (180001)..(195000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (195001)..(210000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (210001)..(225000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (225001)..(240000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (240001)..(255000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (255001)..(270000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (270001)..(285000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (285001)..(300000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (300001)..(315000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (315001)..(330000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (330001)..(345000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (345001)..(360000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (360001)..(375000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (375001)..(390000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (390001)..(405000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (405001)..(420000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (420001)..(435000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (435001)..(450000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (450001)..(465000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (465001)..(480000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (480001)..(495000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (495001)..(510000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (510001)..(525000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (525001)..(540000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (540001)..(555000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (555001)..(570000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (570001)..(585000)
; OTHER INFORMATION: n=a or c or g or t

```
;   NAME/KEY: misc_feature
;   LOCATION: (585001)..(600000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (600001)..(615000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (615001)..(630000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (630001)..(645000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (645001)..(660000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (660001)..(675000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (675001)..(690000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (690001)..(705000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (705001)..(720000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (720001)..(735000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (735001)..(750000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (750001)..(765000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (765001)..(780000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (780001)..(795000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (795001)..(810000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (810001)..(825000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (825001)..(840000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (840001)..(855000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (855001)..(870000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (870001)..(885000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (885001)..(900000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (900001)..(915000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature

  Query Match             0.2%;  Score 20;  DB 4;  Length 1230025;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6172 TTAACACCCTTAACTGGAAT 6191
             ||||||||||||||||||||
Db     907689 TTAACACCCTTAACTGGAAT 907708


RESULT 11
US-09-328-352-2598/c
; Sequence 2598, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACT
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2598
;   LENGTH: 234
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-2598

  Query Match             0.2%;  Score 19;  DB 4;  Length 234;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        263 AAAATTTAAACTGCCTGAT 281
              |||||||||||||||||||
Db         70 AAAATTTAAACTGCCTGAT 52


RESULT 12
US-09-313-294A-4581
; Sequence 4581, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
;  APPLICANT:  Lalgudi, Raghunath V.
;  APPLICANT:  Ito, Laura Y.
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;  FILE REFERENCE:  PL-0017 US
;  CURRENT APPLICATION NUMBER: US/09/313,294A
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS:  7600
;  SOFTWARE:  PERL Program
; SEQ ID NO 4581
;   LENGTH: 284
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6476212 700348665H1
;   NAME/KEY: unsure
;   LOCATION: 44
;   OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-4581

  Query Match             0.2%;  Score 19;  DB 4;  Length 284;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8700 TGGCGGCCGGCTCTCCGCC 8718
              |||||||||||||||||||
Db         22 TGGCGGCCGGCTCTCCGCC 40


RESULT 13
US-09-313-294A-88
; Sequence 88, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
;  APPLICANT:  Lalgudi, Raghunath V.
;  APPLICANT:  Ito, Laura Y.
```

```
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;  FILE REFERENCE:  PL-0017 US
;  CURRENT APPLICATION NUMBER: US/09/313,294A
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS:  7600
;  SOFTWARE:  PERL Program
; SEQ ID NO 88
;   LENGTH: 292
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6476212 700548493H1
;   NAME/KEY: unsure
;   LOCATION: 6, 46, 63-64
;   OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-88

  Query Match             0.2%;  Score 19;  DB 4;  Length 292;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8700 TGGCGGCCGGCTCTCCGCC 8718
             ||||||||||||||||||||
Db         24 TGGCGGCCGGCTCTCCGCC 42


RESULT 14
US-09-328-352-2603
; Sequence 2603, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2603
;   LENGTH: 471
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-2603

  Query Match             0.2%;  Score 19;  DB 4;  Length 471;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        263 AAAATTTAAACTGCCTGAT 281
              ||||||||||||||||||||
Db         92 AAAATTTAAACTGCCTGAT 110


RESULT 15
US-09-620-312D-669/c
; Sequence 669, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Xue, Aidong J.
;  APPLICANT:  Yang, Yonghong
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  John Tillinghast
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. 6569662el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2B
;  CURRENT APPLICATION NUMBER: US/09/620,312D
;  CURRENT FILING DATE:  2000-07-19
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1105
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 669
;   LENGTH: 719
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (81)..(578)
US-09-620-312D-669

  Query Match             0.2%;  Score 19;  DB 4;  Length 719;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2883 CCTGACCAGGCGGGCGAGG 2901
              ||||||||||||||||||||
Db         62 CCTGACCAGGCGGGCGAGG 44


RESULT 16
US-07-885-970A-7
; Sequence 7, Application US/07885970A
; Patent No. 5495070
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E.
;    TITLE OF INVENTION:  GENETICALLY ENGINEERING COTTON
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nicholas J. Seay, Quarles & Brady
;      STREET:  P.O. Box 2113, First Wisconsin Plaza
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/885,970A
;      FILING DATE:  19920518
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/617,239
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/253,243
;      FILING DATE:  04-OCT-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J.
;      REGISTRATION NUMBER:  27,386
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (608) 283-2478
;      TELEFAX:  (608) 251-5139
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  989 base pairs
;      TYPE:  NUCLEIC ACID
```

```
;    STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium hirsutum
;      STRAIN:  Coker 312
;      DEVELOPMENTAL STAGE:  10 day old fiber cells
;      TISSUE TYPE:  fiber cells
;    IMMEDIATE SOURCE:
;      LIBRARY:  CKFB10
;      CLONE:  A11
US-07-885-970A-7

  Query Match             0.2%;  Score 19;  DB 1;  Length 989;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1359 AAATGGAAAAACAAAAAAA 1377
              |||||||||||||||||||
Db        400 AAATGGAAAAACAAAAAAA 418


RESULT 17
US-08-298-687A-7
; Sequence 7, Application US/08298687A
; Patent No. 5521078
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E.
;    TITLE OF INVENTION:  GENETICALLY ENGINEERING COTTON
;    TITLE OF INVENTION:  PLANTS FOR ALTERED FIBER
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nicholas J. Seay, Quarles & Brady
;      STREET:  P.O. Box 2113, First Wisconsin Plaza
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/298,687A
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/617,239
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/253,243
;      FILING DATE:  04-OCT-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J.
;      REGISTRATION NUMBER:  27,386
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (608) 283-2478
;      TELEFAX:  (608) 251-5139
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  989 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium hirsutum
;      STRAIN:  Coker 312
;      DEVELOPMENTAL STAGE:  10 day old fiber cells
;      TISSUE TYPE:  fiber cells
;    IMMEDIATE SOURCE:
;      LIBRARY:  CKFB10
;      CLONE:  A11
US-08-298-687A-7

  Query Match             0.2%;  Score 19;  DB 1;  Length 989;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1359 AAATGGAAAAACAAAAAAA 1377
              |||||||||||||||||||
Db        400 AAATGGAAAAACAAAAAAA 418


RESULT 18
US-08-530-797-6
; Sequence 6, Application US/08530797
; Patent No. 5597718
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E.
;    APPLICANT:  Umbeck, Paul F.
;    APPLICANT:  Brill, Winston J.
;    TITLE OF INVENTION:  GENETICALY ENGINEERED COTTON PLANTS
;    TITLE OF INVENTION:  FOR ALTERED FIBER
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles and Brady
;      STREET:  P.O BOX 2113
;      STREET:  FIRST WISCONSIN PLAZA
;      CITY:  MADISON
;      STATE:  WISCONSIN
;      COUNTRY:  U.S.A.
;      ZIP:  53701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette - 3.50 inch, 800Kb storage
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh
;      SOFTWARE:  Microsoft Word 4.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/530,797
;      FILING DATE:  20-SEP-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/617,239
;      FILING DATE:  21-NOV-90
;      APPLICATION NUMBER:  US 07/253,243
;      FILING DATE:  04-OCT-88
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nicholas J. Seay
;      REGISTRATION NUMBER:  27,386
;      REFERENCE/DOCKET NUMBER:  1122990245
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  989 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium hirsutum
;      STRAIN:  Coker 312
;      DEVELOPMENTAL STAGE:  10 day old fiber cells
;      TISSUE TYPE:  fiber cells
;    IMMEDIATE SOURCE:
;      LIBRARY:  CKFB10
;      CLONE:  A11
US-08-530-797-6

  Query Match             0.2%;  Score 19;  DB 1;  Length 989;
```

```
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1359 AAATGGAAAAACAAAAAAA 1377
             |||||||||||||||||||
Db        400 AAATGGAAAAACAAAAAAA 418


RESULT 19
US-08-298-829-7
; Sequence 7, Application US/08298829
; Patent No. 5620882
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E.
;    TITLE OF INVENTION:  GENETICALLY ENGINEERING COTTON
;    TITLE OF INVENTION:  PLANTS FOR ALTERED FIBER
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nicholas J. Seay, Quarles & Brady
;      STREET:  P.O. Box 2113, First Wisconsin Plaza
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/298,829
;      FILING DATE:  19-OCT-1994
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/885,970
;      FILING DATE:  18-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/617,239
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/253,243
;      FILING DATE:  04-OCT-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J.
;      REGISTRATION NUMBER:  27,386
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (608) 283-2478
;      TELEFAX:  (608) 251-5139
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  989 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium hirsutum
;      STRAIN:  Coker 312
;      DEVELOPMENTAL STAGE:  10 day old fiber cells
;      TISSUE TYPE:  fiber cells
;    IMMEDIATE SOURCE:
;      LIBRARY:  CKFB10
;      CLONE:  A11
US-08-298-829-7

  Query Match             0.2%;  Score 19;  DB 1;  Length 989;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1359 AAATGGAAAAACAAAAAAA 1377
             |||||||||||||||||||
Db        400 AAATGGAAAAACAAAAAAA 418


RESULT 20
US-08-787-335-6
; Sequence 6, Application US/08787335
; Patent No. 5981834
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E.
;    APPLICANT:  Umbeck, Paul F.
;    APPLICANT:  Brill, Winston J.
;    TITLE OF INVENTION:  GENETICALY ENGINEERED COTTON PLANTS
;    TITLE OF INVENTION:  FOR ALTERED FIBER
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles and Brady
;      STREET:  P.O BOX 2113
;      STREET:  FIRST WISCONSIN PLAZA
;      CITY:  MADISON
;      STATE:  WISCONSIN
;      COUNTRY:  U.S.A.
;      ZIP:  53701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette - 3.50 inch, 800Kb storage
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh
;      SOFTWARE:  Microsoft Word 4.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/787,335
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/530,797
;      FILING DATE:
;      APPLICATION NUMBER:  US 07/253,243
;      FILING DATE:  04-OCT-88
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nicholas J. Seay
;      REGISTRATION NUMBER:  27,386
;      REFERENCE/DOCKET NUMBER:  1122990245
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  989 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium hirsutum
;      STRAIN:  Coker 312
;      DEVELOPMENTAL STAGE:  10 day old fiber cells
;      TISSUE TYPE:  fiber cells
;    IMMEDIATE SOURCE:
;      LIBRARY:  CKFB10
;      CLONE:  A11
US-08-787-335-6

  Query Match             0.2%;  Score 19;  DB 2;  Length 989;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1359 AAATGGAAAAACAAAAAAA 1377
             |||||||||||||||||||
Db        400 AAATGGAAAAACAAAAAAA 418


RESULT 21
US-09-328-352-49
; Sequence 49, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
```

```
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 49
;   LENGTH: 1215
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-49

  Query Match             0.2%;  Score 19;  DB 4;  Length 1215;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8176 TAATGTCTTAGAAAATCCA 8194
             ||||||||||||||||||||
Db        726 TAATGTCTTAGAAAATCCA 744


RESULT 22
US-08-299-849B-21/c
; Sequence 21, Application US/08299849B
; Patent No. 5612201
;  GENERAL INFORMATION:
;    APPLICANT:  De Plaen, Etienne; Boon-Falleur, Thierry;
;    APPLICANT:  Leth , Bernard; Szikora, Jean-Pierre; De Smet, Charles;
;    APPLICANT:  Chomez, Patrick
;    TITLE OF INVENTION:  Isolated Nucleic Acid Molecules Useful In
;    TITLE OF INVENTION:  Determining Expression Of A Tumor Antigen Precursor
;    NUMBER OF SEQUENCES:  48
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/299,849B
;      FILING DATE:  1-SEPTEMBER-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/037,230
;      FILING DATE:  26-MARCH-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/04354
;      FILING DATE:  22-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,043
;      FILING DATE:  12-DECEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/764,364
;      FILING DATE:  23-SEPTEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/728,838
;      APPLICATION NUMBER:  9-JULY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/705,702
;      FILING DATE:  23-May-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5612201man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5355
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1412 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      NAME/KEY:  MAGE-9 gene
US-08-299-849B-21

  Query Match             0.2%;  Score 19;  DB 1;  Length 1412;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3426 GGAGTGGCCGGCGGGGTCC 3444
              ||||||||||||||||||||
Db        949 GGAGTGGCCGGCGGGGTCC 931


RESULT 23
US-08-142-368A-21/c
; Sequence 21, Application US/08142368A
; Patent No. 5925729
;  GENERAL INFORMATION:
;    APPLICANT:  Boon-Falleur, Thierry; Van der Bruggen, Thierry;
;    APPLICANT:  Van den Eynde, Beno t; Van Pel, Aline; De Plaen, Etienne;
;    APPLICANT:  Lurquin, Christophe; Chomez, Patrick; Traversari, Catia
;    TITLE OF INVENTION:  Tumor Rejection Antigen Precursors, Tumor
;    TITLE OF INVENTION:  Rejection Antigens and Uses Thereof
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/142,368A
;      FILING DATE:  02-MAY-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/04354
;      FILING DATE:  22-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,043
;      FILING DATE:  12-DECEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/764,364
;      FILING DATE:  23-SEPTEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/728,838
;      APPLICATION NUMBER:  9-JULY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/705,702
;      FILING DATE:  23-May-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5925729man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5253.4-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1412 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      NAME/KEY:  MAGE-9 gene
US-08-142-368A-21

  Query Match             0.2%;  Score 19;  DB 2;  Length 1412;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3426 GGAGTGGCCGGCGGGGTCC 3444
              ||||||||||||||||||||
Db        949 GGAGTGGCCGGCGGGGTCC 931


RESULT 24
US-08-967-727-21/c
; Sequence 21, Application US/08967727
; Patent No. 6025474
;  GENERAL INFORMATION:
;    APPLICANT:  Gaugler, B atrice; Van den Eynde, Beno t;
;    APPLICANT:  van der Bruggen, Pierre;  Boon-Falleur, Thierry
;    TITLE OF INVENTION:  Isolated Nucleic Acid Molecules Coding For
;    TITLE OF INVENTION:  Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/967,727
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/037,230
;      FILING DATE:  26-MARCH-1993
;      APPLICATION NUMBER:  PCT/US92/04354
;      FILING DATE:  22-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,043
;      FILING DATE:  12-DECEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/764,365
;      FILING DATE:  23-SEPTEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/728,838
;      FILING DATE:  9-JULY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/705,702
;      FILING DATE:  23-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 6025474man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5353
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1412 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      NAME/KEY:  MAGE-9 gene
US-08-967-727-21

  Query Match             0.2%;  Score 19;  DB 3;  Length 1412;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3426 GGAGTGGCCGGCGGGGTCC 3444
              ||||||||||||||||||||
Db        949 GGAGTGGCCGGCGGGGTCC 931


RESULT 25
US-08-037-230D-21/c
; Sequence 21, Application US/08037230D
; Patent No. 6235525
;  GENERAL INFORMATION:
;    APPLICANT:  Gaugler, B atrice; Van den Eynde, Beno t;
;    APPLICANT:  van der Bruggen, Pierre;  Boon-Falleur, Thierry
;    TITLE OF INVENTION:  Isolated Nucleic Acid Molecules Coding For
;    TITLE OF INVENTION:  Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/037,230D
;      FILING DATE:  26-MARCH-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/04354
;      FILING DATE:  22-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,043
;      FILING DATE:  12-DECEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/764,364
;      FILING DATE:  23-SEPTEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/728,838
;      FILING DATE:  9-JULY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/705,702
;      FILING DATE:  23-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 6235525man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5353
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1412 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      NAME/KEY:  MAGE-9 gene
US-08-037-230D-21

  Query Match             0.2%;  Score 19;  DB 3;  Length 1412;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       3426 GGAGTGGCCGGCGGGGTCC 3444
              ||||||||||||||||||
Db        949 GGAGTGGCCGGCGGGGTCC 931


RESULT 26
US-09-583-850-21/c
; Sequence 21, Application US/09583850
; Patent No. 6498021
;  GENERAL INFORMATION:
;    APPLICANT:  Gaugler, Batrice; Van den Eynde, BenoEt;
;    APPLICANT:  van der Bruggen, Pierre;  Boon-Falleur, Thierry
;    TITLE OF INVENTION:  Isolated Nucleic Acid Molecules Coding For
;    TITLE OF INVENTION:  Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/583,850
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/583,613
;      FILING DATE:
;      APPLICATION NUMBER:  PCT/US92/04354
;      FILING DATE:  22-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,043
;      FILING DATE:  12-DECEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/764,365
;      FILING DATE:  23-SEPTEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/728,838
;      FILING DATE:  9-JULY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/705,702
;      FILING DATE:  23-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 6498021man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5353
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1412 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      NAME/KEY:  MAGE-9 gene
US-09-583-850-21

  Query Match             0.2%;  Score 19;  DB 4;  Length 1412;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3426 GGAGTGGCCGGCGGGGTCC 3444
              ||||||||||||||||||
Db        949 GGAGTGGCCGGCGGGGTCC 931


RESULT 27
US-09-579-197-21/c
; Sequence 21, Application US/09579197
; Patent No. 6552180
;  GENERAL INFORMATION:
;    APPLICANT:  Gaugler, Beatrice; Van den Eynde, Benoit;
;    APPLICANT:  van der Bruggen, Pierre;  Boon-Falleur, Thierry
;    TITLE OF INVENTION:  Isolated Nucleic Acid Molecules Coding
;    TITLE OF INVENTION:  For Tumor Rejection Antigen Precursor Mage-3 And Uses Ther
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/579,197
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/037,230
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,043
;      FILING DATE:  12-DECEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/764,364
;      FILING DATE:  23-SEPTEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/728,838
;      FILING DATE:  9-JULY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/705,702
;      FILING DATE:  23-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 6552180man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5353
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1412 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      NAME/KEY:  MAGE-9 gene
US-09-579-197-21

  Query Match             0.2%;  Score 19;  DB 4;  Length 1412;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3426 GGAGTGGCCGGCGGGGTCC 3444
              ||||||||||||||||||
Db        949 GGAGTGGCCGGCGGGGTCC 931


RESULT 28
US-09-404-026-21/c
; Sequence 21, Application US/09404026
; Patent No. 6565857
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Gaugler, B atrice; Van den Eynde, Beno t;
;    APPLICANT:  van der Bruggen, Pierre;  Boon-Falleur, Thierry
;    TITLE OF INVENTION:  Isolated Nucleic Acid Molecules Coding For
;    TITLE OF INVENTION:  Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/404,026
;      FILING DATE:  23-SEPT-1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/037,230
;      FILING DATE:  26-MARCH-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/04354
;      FILING DATE:  22-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,043
;      FILING DATE:  12-DECEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/764,364
;      FILING DATE:  23-SEPTEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/728,838
;      FILING DATE:  9-JULY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/705,702
;      FILING DATE:  23-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 6565857man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5353
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1412 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      NAME/KEY:  MAGE-9 gene
US-09-404-026-21

  Query Match             0.2%;  Score 19;  DB 4;  Length 1412;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3426 GGAGTGGCCGGCGGGGTCC 3444
              |||||||||||||||||||
Db        949 GGAGTGGCCGGCGGGGTCC 931


RESULT 29
US-08-979-847B-117/c
; Sequence 117, Application US/08979847B
; Patent No. 6582703
;   GENERAL INFORMATION:
;        APPLICANT: PERRON, HERVE
;                   BESEME, FREDERIC
;                   BEDIN, FREDERIC
;                   PARANHOS-BACCALA, GLAUCIA
;                   KOMURIAN-PRADEL, FLORENCE
;                   JOLIVET-REYNAUD, COLETTE
;                   MANDRAND, BERNARD
;                   GARSON, JEREMY
;                   TUKE, PHILIP
;        TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                            ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROP
;                            THERAPEUTIC PURPOSES
;        NUMBER OF SEQUENCES: 210
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: OLIFF & BERRIDGE, PLC
;             STREET: P.O. BOX 19928
;             CITY: ALEXANDRIA
;             STATE: VA
;             COUNTRY: USA
;             ZIP: 22320
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/979,847B
;             FILING DATE: 26-No. 6582703-1997
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: BERRIDGE, WILLIAM P.
;             REGISTRATION NUMBER: 30,024
;             REFERENCE/DOCKET NUMBER: WPB 39046A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 703-836-6400
;             TELEFAX: 703-836-2787
;   INFORMATION FOR SEQ ID NO: 117:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1511 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 117:
US-08-979-847B-117

  Query Match             0.2%;  Score 19;  DB 4;  Length 1511;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        377 ACATCAATATAGCCGTCAG 395
              |||||||||||||||||||
Db        558 ACATCAATATAGCCGTCAG 540


RESULT 30
US-08-461-775-9
; Sequence 9, Application US/08461775
; Patent No. 5858773
;  GENERAL INFORMATION:
;    APPLICANT:  MAZODIER, Philippe
;    APPLICANT:  GUGLIEMI, Gerard
;    TITLE OF INVENTION:  REGULATORY NUCLEOTIDE SEQUENCE OF THE
;    TITLE OF INVENTION:  INITIATION OF TRANSCRIPTION
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  George Mason Bldg., Washington & Prince Sts.
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;    SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/461,775
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/050,313
;      FILING DATE:  10-MAY-1993
;      APPLICATION NUMBER:  FR 9011186
;      FILING DATE:  10-SEP-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crane-Feury, Sharon E
;      REGISTRATION NUMBER:  36,113
;      REFERENCE/DOCKET NUMBER:  010830-035
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2167 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-461-775-9

  Query Match             0.2%;  Score 19;  DB 2;  Length 2167;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6910 GGGGGAGGTCGGATCGTGA 6928
             ||||||||||||||||||||
Db        368 GGGGGAGGTCGGATCGTGA 386


RESULT 31
US-09-031-606-9
; Sequence 9, Application US/09031606
; Patent No. 6153404
;  GENERAL INFORMATION:
;    APPLICANT:  MAZODIER, Philippe
;    APPLICANT:  GUGLIEMI, Gerard
;    TITLE OF INVENTION:  REGULATORY NUCLEOTIDE SEQUENCE OF THE
;    TITLE OF INVENTION:  INITIATION OF TRANSCRIPTION
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  George Mason Bldg., Washington & Prince Sts.
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/031,606
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/050,313
;      FILING DATE:  10-MAY-1993
;      APPLICATION NUMBER:  FR 9011186
;      FILING DATE:  10-SEP-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crane-Feury, Sharon E
;      REGISTRATION NUMBER:  36,113
;      REFERENCE/DOCKET NUMBER:  010830-035
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2167 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-031-606-9

  Query Match             0.2%;  Score 19;  DB 3;  Length 2167;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6910 GGGGGAGGTCGGATCGTGA 6928
             ||||||||||||||||||||
Db        368 GGGGGAGGTCGGATCGTGA 386


RESULT 32
US-08-461-775-11
; Sequence 11, Application US/08461775
; Patent No. 5858773
;  GENERAL INFORMATION:
;    APPLICANT:  MAZODIER, Philippe
;    APPLICANT:  GUGLIEMI, Gerard
;    TITLE OF INVENTION:  REGULATORY NUCLEOTIDE SEQUENCE OF THE
;    TITLE OF INVENTION:  INITIATION OF TRANSCRIPTION
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  George Mason Bldg., Washington & Prince Sts.
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/461,775
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/050,313
;      FILING DATE:  10-MAY-1993
;      APPLICATION NUMBER:  FR 9011186
;      FILING DATE:  10-SEP-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crane-Feury, Sharon E
;      REGISTRATION NUMBER:  36,113
;      REFERENCE/DOCKET NUMBER:  010830-035
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2668 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-461-775-11

  Query Match             0.2%;  Score 19;  DB 2;  Length 2668;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6910 GGGGGAGGTCGGATCGTGA 6928
             ||||||||||||||||||||
Db        368 GGGGGAGGTCGGATCGTGA 386
```

```
RESULT 33
US-09-031-606-11
; Sequence 11, Application US/09031606
; Patent No. 6153404
;  GENERAL INFORMATION:
;    APPLICANT:  MAZODIER, Philippe
;    APPLICANT:  GUGLIEMI, Gerard
;    TITLE OF INVENTION:  REGULATORY NUCLEOTIDE SEQUENCE OF THE
;    TITLE OF INVENTION:  INITIATION OF TRANSCRIPTION
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  George Mason Bldg., Washington & Prince Sts.
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/031,606
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/050,313
;      FILING DATE:  10-MAY-1993
;      APPLICATION NUMBER:  FR 9011186
;      FILING DATE:  10-SEP-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crane-Feury, Sharon E
;      REGISTRATION NUMBER:  36,113
;      REFERENCE/DOCKET NUMBER:  010830-035
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2668 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-031-606-11

  Query Match             0.2%;  Score 19;  DB 3;  Length 2668;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6910 GGGGGAGGTCGGATCGTGA 6928
             ||||||||||||||||||||
Db        368 GGGGGAGGTCGGATCGTGA 386


RESULT 34
US-09-056-105-15/c
; Sequence 15, Application US/09056105
; Patent No. 6287569
; GENERAL INFORMATION:
;  APPLICANT: KIPPS, THOMAS J.
;  APPLICANT: WU, YUNQI
;  TITLE OF INVENTION: VACCINES WITH ENHANCED INTRACELLULAR
;  TITLE OF INVENTION: PROCESSING
;  FILE REFERENCE: 233/221
;  CURRENT APPLICATION NUMBER: US/09/056,105
;  CURRENT FILING DATE: 1998-04-06
;  EARLIER APPLICATION NUMBER: 60/043,467
;  EARLIER FILING DATE: 1997-04-10
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 15
;   LENGTH: 2931
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-056-105-15

  Query Match             0.2%;  Score 19;  DB 3;  Length 2931;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3426 GGAGTGGCCGGCGGGGTCC 3444
             ||||||||||||||||||||
Db       1854 GGAGTGGCCGGCGGGGTCC 1836


RESULT 35
US-09-300-008B-1
; Sequence 1, Application US/09300008B
; Patent No. 6458534
; GENERAL INFORMATION:
;  APPLICANT: Concannon et al.
;  TITLE OF INVENTION: A GENE ASSOCIATED WITH NIJMEGEN BREAKAGE
;  TITLE OF INVENTION:  SYNDROME, ITS GENE PRODUCT AND METHODS FOR THEIR USE
;  FILE REFERENCE: 9924-0003-228
;  CURRENT APPLICATION NUMBER: US/09/300,008B
;  CURRENT FILING DATE:  1999-04-27
;  PRIOR APPLICATION NUMBER: US 60/083,269
;  PRIOR FILING DATE: 1998-04-27
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 4386
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (26)...(2287)
US-09-300-008B-1

  Query Match             0.2%;  Score 19;  DB 4;  Length 4386;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3375 AAACGTCTTGTTCTCTTTT 3393
             ||||||||||||||||||||
Db       1476 AAACGTCTTGTTCTCTTTT 1494


RESULT 36
US-09-177-419C-11/c
; Sequence 11, Application US/09177419C
; Patent No. 6562609
; GENERAL INFORMATION:
;  APPLICANT: Russel, David W
;  APPLICANT:  Lund, Erik G
;  TITLE OF INVENTION: Cholesterol 25-Hydroxylase
;  FILE REFERENCE: UTSD1370
;  CURRENT APPLICATION NUMBER: US/09/177,419C
;  CURRENT FILING DATE:  1998-10-22
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 36
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Sequence
US-09-177-419C-11

  Query Match             0.2%;  Score 18;  DB 4;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       5229 ATGGTGAGTGGGTGACAC 5246
              ||||||||||||||||||
Db         33 ATGGTGAGTGGGTGACAC 16


RESULT 37
US-09-134-001C-1986/c
; Sequence 1986, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1986
;   LENGTH: 645
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1986

  Query Match             0.2%;  Score 18;  DB 4;  Length 645;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4450 AACAATTGAAGCTCTTCT 4467
              ||||||||||||||||||
Db         68 AACAATTGAAGCTCTTCT 51


RESULT 38
US-09-134-001C-987
; Sequence 987, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 987
;   LENGTH: 696
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-987

  Query Match             0.2%;  Score 18;  DB 4;  Length 696;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        225 TGGTGCTGGGAAAAGTAC 242
              ||||||||||||||||||
Db        129 TGGTGCTGGGAAAAGTAC 146


RESULT 39
US-08-203-905B-3
; Sequence 3, Application US/08203905B
; Patent No. 5646249
;  GENERAL INFORMATION:
;    APPLICANT:  KAYE, FEDERIC J.
;    APPLICANT:  OTTERSON, GREGORY A.
;    TITLE OF INVENTION:  ISOLATION AND CHARACTERIZATION OF A
;    TITLE OF INVENTION:  NOVEL CHAPERONE PROTEIN
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KNOBBE, MARTENS, OLSON & BEAR
;      STREET:  620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
;      CITY:  NEWPORT BEACH
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/203,905B
;      FILING DATE:  February 28, 1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KIRKPATRICK, ANITA M.
;      REGISTRATION NUMBER:  32,617
;      REFERENCE/DOCKET NUMBER:  NIH089.001A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-235-8550
;      TELEFAX:  619-235-0176
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  792 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-203-905B-3

  Query Match             0.2%;  Score 18;  DB 1;  Length 792;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4445 AATTCAACAATTGAAGCT 4462
              ||||||||||||||||||
Db        223 AATTCAACAATTGAAGCT 240


RESULT 40
US-08-467-963C-28
; Sequence 28, Application US/08467963C
; Patent No. 5968776
;  GENERAL INFORMATION:
;    APPLICANT:  KLEIN, Michel H
;    APPLICANT:  DU, Run-Pan
;    APPLICANT:  EWASYSHYN, Mary E
;    TITLE OF INVENTION:  MULTIMERIC HYBRID GENE ENCODING A
;    TITLE OF INVENTION:  CHIMERIC PROTEIN WHICH CONFERS PROTECTION AGAINST
;    TITLE OF INVENTION:  PARAINFLUENZA VIRUS AND RESPIRATORY SYNCYTIAL VIRUS
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  6th Floor, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/467,963C
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/838,189
;      FILING DATE:  16-APR-1997
;      APPLICATION NUMBER:  US 08/001,554
;      FILING DATE:  06-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9200117.1
;      FILING DATE:  06-JAN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEWART, Michael I
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-474 MIS:jb
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  894 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-467-963C-28

  Query Match             0.2%;  Score 18;  DB 2;  Length 894;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2345 ACAACCCAAACACAACCC 2362
              ||||||||||||||||||
Db        412 ACAACCCAAACACAACCC 429


RESULT 41
US-08-838-189D-28
; Sequence 28, Application US/08838189D
; Patent No. 5998169
;  GENERAL INFORMATION:
;    APPLICANT:  KLEIN, Michel H
;    APPLICANT:  DU, Run-Pan
;    APPLICANT:  EWASYSHYN, Mary E
;    TITLE OF INVENTION:  MULTIMERIC HYBRID GENE ENCODING A
;    TITLE OF INVENTION:  CHIMERIC PROTEIN WHICH CONFERS PROTECTION AGAINST
;    TITLE OF INVENTION:  PARAINFLUENZA VIRUS AND RESPIRATORY SYNCYTIAL VIRUS
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  6th Floor, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/838,189D
;      FILING DATE:  16-APR-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/001,554
;      FILING DATE:  06-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9200117.1
;      FILING DATE:  06-JAN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEWART, Michael I
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-687 MIS:jb
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  894 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-838-189D-28

  Query Match             0.2%;  Score 18;  DB 2;  Length 894;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2345 ACAACCCAAACACAACCC 2362
              ||||||||||||||||||
Db        412 ACAACCCAAACACAACCC 429


RESULT 42
US-08-852-344D-28
; Sequence 28, Application US/08852344D
; Patent No. 6017539
;  GENERAL INFORMATION:
;    APPLICANT:  KLEIN, Michel H
;    APPLICANT:  DU, Run-Pan
;    APPLICANT:  EWASYSHYN, Mary E
;    TITLE OF INVENTION:  CHIMERIC PROTEIN WHICH CONFERS PROTECTION
;    TITLE OF INVENTION:  AGAINST PARAINFLUENZA VIRUS AND RESPIRATORY
;    TITLE OF INVENTION:  SYNCYTIAL VIRUS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  6th Floor, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/852,344D
;      FILING DATE:  07-MAY-1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/344,639
;      FILING DATE:  14-NOV-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9200117.1
;      FILING DATE:  06-JAN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEWART, Michael I
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-688 MIS:jb
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  894 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-852-344D-28

  Query Match             0.2%;  Score 18;  DB 3;  Length 894;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      2345 ACAACCCAAACACAACCC 2362
             ||||||||||||||||||
Db       412 ACAACCCAAACACAACCC 429


RESULT 43
US-08-344-639E-28
; Sequence 28, Application US/08344639E
; Patent No. 6033668
;  GENERAL INFORMATION:
;    APPLICANT:  Klein, Michel H
;    APPLICANT:  Du, Run-Pan
;    APPLICANT:  Ewasyshyn, Mary E
;    TITLE OF INVENTION:  CHIMERIC PROTEIN WHICH CONFERS
;    TITLE OF INVENTION:  PROTECTION AGAINST PARAINFLUENZA VIRUS
;    TITLE OF INVENTION:  AND RESPIRATORY SYNCYTIAL VIRUS
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  330 University Avenue, 6th Floor
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/344,639E
;      FILING DATE:  14-NOV-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/001,554
;      FILING DATE:  06-JAN-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9200117.1
;      FILING DATE:  06-JAN-1992
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stewart, Michael I
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-391 MIS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;      TELEX:  065-24567 SIMBAS
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  894 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-344-639E-28

  Query Match             0.2%;  Score 18;  DB 3;  Length 894;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2345 ACAACCCAAACACAACCC 2362
             ||||||||||||||||||
Db       412 ACAACCCAAACACAACCC 429


RESULT 44
US-08-467-963C-7
; Sequence 7, Application US/08467963C
; Patent No. 5968776
;  GENERAL INFORMATION:
;    APPLICANT:  KLEIN, Michel H
;    APPLICANT:  DU, Run-Pan
;    APPLICANT:  EWASYSHYN, Mary E
;    TITLE OF INVENTION:  MULTIMERIC HYBRID GENE ENCODING A
;    TITLE OF INVENTION:  CHIMERIC PROTEIN WHICH CONFERS PROTECTION AGAINST
;    TITLE OF INVENTION:  PARAINFLUENZA VIRUS AND RESPIRATORY SYNCYTIAL VIRUS
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  6th Floor, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/467,963C
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/838,189
;      FILING DATE:  16-APR-1997
;      APPLICATION NUMBER:  US 08/001,554
;      FILING DATE:  06-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9200117.1
;      FILING DATE:  06-JAN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEWART, Michael I
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-474 MIS:jb
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  920 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-467-963C-7

  Query Match             0.2%;  Score 18;  DB 2;  Length 920;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2345 ACAACCCAAACACAACCC 2362
             ||||||||||||||||||
Db       419 ACAACCCAAACACAACCC 436


RESULT 45
US-08-838-189D-7
; Sequence 7, Application US/08838189D
; Patent No. 5998169
;  GENERAL INFORMATION:
;    APPLICANT:  KLEIN, Michel H
;    APPLICANT:  DU, Run-Pan
;    APPLICANT:  EWASYSHYN, Mary E
;    TITLE OF INVENTION:  MULTIMERIC HYBRID GENE ENCODING A
;    TITLE OF INVENTION:  CHIMERIC PROTEIN WHICH CONFERS PROTECTION AGAINST
;    TITLE OF INVENTION:  PARAINFLUENZA VIRUS AND RESPIRATORY SYNCYTIAL VIRUS
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  6th Floor, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/838,189D
;      FILING DATE:  16-APR-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/001,554
;      FILING DATE:  06-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9200117.1
;      FILING DATE:  06-JAN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEWART, Michael I
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-687 MIS:jb
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  920 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-838-189D-7

  Query Match             0.2%;  Score 18;  DB 2;  Length 920;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2345 ACAACCCAAACACAACCC 2362
              ||||||||||||||||||
Db        419 ACAACCCAAACACAACCC 436


RESULT 46
US-08-852-344D-7
; Sequence 7, Application US/08852344D
; Patent No. 6017539
;  GENERAL INFORMATION:
;    APPLICANT:  KLEIN, Michel H
;    APPLICANT:  DU, Run-Pan
;    APPLICANT:  EWASYSHYN, Mary E
;    TITLE OF INVENTION:  CHIMERIC PROTEIN WHICH CONFERS PROTECTION
;    TITLE OF INVENTION:  AGAINST PARAINFLUENZA VIRUS AND RESPIRATORY
;    TITLE OF INVENTION:  SYNCYTIAL VIRUS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  6th Floor, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/852,344D
;      FILING DATE:  07-MAY-1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/344,639
;      FILING DATE:  14-NOV-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9200117.1
;      FILING DATE:  06-JAN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEWART, Michael I
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-688 MIS:jb
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  920 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-852-344D-7

  Query Match             0.2%;  Score 18;  DB 3;  Length 920;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2345 ACAACCCAAACACAACCC 2362
              ||||||||||||||||||
Db        419 ACAACCCAAACACAACCC 436


RESULT 47
US-08-344-639E-7
; Sequence 7, Application US/08344639E
; Patent No. 6033668
;  GENERAL INFORMATION:
;    APPLICANT:  Klein, Michel H
;    APPLICANT:  Du, Run-Pan
;    APPLICANT:  Ewasyshyn, Mary E
;    TITLE OF INVENTION:  CHIMERIC PROTEIN WHICH CONFERS
;    TITLE OF INVENTION:  PROTECTION AGAINST PARAINFLUENZA VIRUS
;    TITLE OF INVENTION:  AND RESPIRATORY SYNCYTIAL VIRUS
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  330 University Avenue, 6th Floor
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/344,639E
;      FILING DATE:  14-NOV-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/001,554
;      FILING DATE:  06-JAN-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9200117.1
;      FILING DATE:  06-JAN-1992
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stewart, Michael I
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-391 MIS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;      TELEX:  065-24567 SIMBAS
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  920 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  DNA (genomic)
US-08-344-639E-7

  Query Match             0.2%;  Score 18;  DB 3;  Length 920;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2345 ACAACCCAAACACAACCC 2362
              ||||||||||||||||||
Db        419 ACAACCCAAACACAACCC 436


RESULT 48
US-08-467-969A-7
; Sequence 7, Application US/08467969A
; Patent No. 6168786
;  GENERAL INFORMATION:
;    APPLICANT:  Klein, Michel H
;    APPLICANT:  Du, Run-Pan
;    APPLICANT:  Ewasyshyn, Mary E
;    TITLE OF INVENTION:  Chimeric Immunogens
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  330 University Avenue, 6th Floor
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/467,969A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/001,554
;      FILING DATE:  06-JAN-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9200117.1
;      FILING DATE:  06-JAN-1992
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stewart, Michael I
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-475 MIS:bh
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;      TELEX:  065-24567 SIMBAS
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  920 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-467-969A-7

  Query Match             0.2%;  Score 18;  DB 3;  Length 920;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2345 ACAACCCAAACACAACCC 2362
              ||||||||||||||||||
Db        419 ACAACCCAAACACAACCC 436


RESULT 49
US-08-467-961A-7
; Sequence 7, Application US/08467961A
; Patent No. 6171783
;  GENERAL INFORMATION:
;    APPLICANT:  Klein, Michel H
;    APPLICANT:  Du, Run-Pan
;    APPLICANT:  Ewasyshyn, Mary E
;    TITLE OF INVENTION:  Chimeric Immunogens
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  330 University Avenue, 6TH Floor
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/467,961A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/001,554
;      FILING DATE:  06-JAN-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9200117.1
;      FILING DATE:  06-JAN-1992
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stewart, Michael I
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-476 MIS:bh
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;      TELEX:  065-24567 SIMBAS
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  920 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-467-961A-7

  Query Match             0.2%;  Score 18;  DB 3;  Length 920;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2345 ACAACCCAAACACAACCC 2362
              ||||||||||||||||||
Db        419 ACAACCCAAACACAACCC 436


RESULT 50
US-08-001-554A-7
; Sequence 7, Application US/08001554A
; Patent No. 6225091
;  GENERAL INFORMATION:
;    APPLICANT:  Klein, Michel H
;    APPLICANT:  Du, Run-Pan
;    APPLICANT:  Ewasyshyn, Mary E
;    TITLE OF INVENTION:  Chimeric Immunogens
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  330 University Avenue, 6th Floor
;      CITY:  Toronto
```

```
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/001,554A
;      FILING DATE:  06-JAN-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stewart, Michael I
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-286
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;      TELEX:  065-24567 SIMBAS
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  920 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-001-554A-7

  Query Match             0.2%;  Score 18;  DB 3;  Length 920;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2345 ACAACCCAAACACAACCC 2362
              ||||||||||||||||||
Db        419 ACAACCCAAACACAACCC 436


RESULT 51
US-09-057-860A-3/c
; Sequence 3, Application US/09057860A
; Patent No. 6277820
;  GENERAL INFORMATION:
;    APPLICANT:  Arnon Rosenthal
;    APPLICANT:  Mary Hynes
;    APPLICANT:  Weilan Ye
;    TITLE OF INVENTION:  Method Of Dopaminergic And Seratonergic
;    TITLE OF INVENTION:  Neuron Formation From Neuroprogenitor Cells
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/057,860A
;      FILING DATE:  09-Apr-1998
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1364
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  997 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Circular
US-09-057-860A-3

  Query Match             0.2%;  Score 18;  DB 3;  Length 997;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3071 GCTCGGTGGTGTGGTGGC 3088
              ||||||||||||||||||
Db        732 GCTCGGTGGTGTGGTGGC 715


RESULT 52
US-09-252-991A-1942/c
; Sequence 1942, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 1942
;   LENGTH: 1239
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1942

  Query Match             0.2%;  Score 18;  DB 4;  Length 1239;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3062 CGCGGGCGCGCTCGGTGG 3079
              ||||||||||||||||||
Db       1168 CGCGGGCGCGCTCGGTGG 1151


RESULT 53
US-09-149-476-230
; Sequence 230, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 186 Human Secreted proteins
;  FILE REFERENCE: PZ002P1
;  CURRENT APPLICATION NUMBER: US/09/149,476
;  CURRENT FILING DATE: 1998-09-08
;  EARLIER APPLICATION NUMBER: PCT/US98/04493
;  EARLIER FILING DATE: 1998-03-06
;  EARLIER APPLICATION NUMBER: 60/040,162
;  EARLIER FILING DATE: 1997-03-07
;  EARLIER APPLICATION NUMBER: 60/040,333
;  EARLIER FILING DATE: 1997-03-07
;  EARLIER APPLICATION NUMBER: 60/038,621
;  EARLIER FILING DATE: 1997-03-07
;  EARLIER APPLICATION NUMBER: 60/040,626
;  EARLIER FILING DATE: 1997-03-07
;  EARLIER APPLICATION NUMBER: 60/040,334
;  EARLIER FILING DATE: 1997-03-07
;  EARLIER APPLICATION NUMBER: 60/040,336
;  EARLIER FILING DATE: 1997-03-07
;  EARLIER APPLICATION NUMBER: 60/040,163
;  EARLIER FILING DATE: 1997-03-07
```

; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,879
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,880
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,894
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,911
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,636
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,874
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,910
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,864
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,631
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,845
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,892
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,761
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/047,595
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,599
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,588
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,585
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,586
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,590
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,594
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,589
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,593
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,614
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,578
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,576
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/047,501
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,670
; EARLIER FILING DATE: 1997-04-11

```
;  EARLIER APPLICATION NUMBER: 60/056,632
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,664
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,876
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,881
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,909
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,875
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,862
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,887
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,908
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/048,964
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/057,650
;  EARLIER FILING DATE: 1997-09-05
;  EARLIER APPLICATION NUMBER: 60/056,884
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/057,669
;  EARLIER FILING DATE: 1997-09-05
;  EARLIER APPLICATION NUMBER: 60/049,610
;  EARLIER FILING DATE: 1997-06-13
;  EARLIER APPLICATION NUMBER: 60/061,060
;  EARLIER FILING DATE: 1997-10-02

  Query Match             0.2%;  Score 18;  DB 4;  Length 1336;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3192 TCTTCTTCATCCACATCA 3209
              ||||||||||||||||||
Db        980 TCTTCTTCATCCACATCA 997


RESULT 54
US-08-936-165A-67
; Sequence 67, Application US/08936165A
; Patent No. 6348582
;  GENERAL INFORMATION:
;    APPLICANT:  Black, Michael
;    APPLICANT:  Burnham, Martin
;    APPLICANT:  Hodgson, John
;    APPLICANT:  Knowles, David
;    APPLICANT:  Lonetto, Michael
;    APPLICANT:  Nicholas, Richard
;    APPLICANT:  Pratt, Julie
;    APPLICANT:  Reichard, Richard
;    APPLICANT:  Rosenberg, Martin
;    APPLICANT:  Ward, Judith
;    TITLE OF INVENTION:  No. 6348582el Prokaryotic Polynucleotides,
;    TITLE OF INVENTION:  Polypeptides and Their Uses
;    NUMBER OF SEQUENCES:  534
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/936,165A
;      FILING DATE:  24-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/027,032
;      FILING DATE:  24-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gimmi, Edward R
;      REGISTRATION NUMBER:  38,891
;      REFERENCE/DOCKET NUMBER:  P50549
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-4478
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  67:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1442 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-08-936-165A-67

  Query Match             0.2%;  Score 18;  DB 4;  Length 1442;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2376 AAATAAAAACAAGCAAAT 2393
              ||||||||||||||||||
Db        924 AAATAAAAACAAGCAAAT 941


RESULT 55
US-09-149-476-65/c
; Sequence 65, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 186 Human Secreted proteins
;  FILE REFERENCE: PZ002P1
;  CURRENT APPLICATION NUMBER: US/09/149,476
;  CURRENT FILING DATE: 1998-09-08
;  EARLIER APPLICATION NUMBER: PCT/US98/04493
;  EARLIER FILING DATE: 1998-03-06
;  EARLIER APPLICATION NUMBER: 60/040,162
;  EARLIER FILING DATE: 1997-03-07
;  EARLIER APPLICATION NUMBER: 60/040,333
;  EARLIER FILING DATE: 1997-03-07
;  EARLIER APPLICATION NUMBER: 60/038,621
;  EARLIER FILING DATE: 1997-03-07
;  EARLIER APPLICATION NUMBER: 60/040,626
;  EARLIER FILING DATE: 1997-03-07
;  EARLIER APPLICATION NUMBER: 60/040,334
;  EARLIER FILING DATE: 1997-03-07
;  EARLIER APPLICATION NUMBER: 60/040,336
;  EARLIER FILING DATE: 1997-03-07
;  EARLIER APPLICATION NUMBER: 60/040,163
;  EARLIER FILING DATE: 1997-03-07
;  EARLIER APPLICATION NUMBER: 60/047,600
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,615
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,597
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,502
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,633
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,583
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,617
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,618
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,503
```

; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,879
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,880
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,894
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,911
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,636
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,874
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,910
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,864
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,631
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,845
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,892
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,761
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/047,595
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,599
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,588
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,585
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,586
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,590
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,594
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,589
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,593
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,614
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,578
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,576
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/047,501
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,670
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/056,632
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,664
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,876
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,881
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,909
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,875
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,862
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,887
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,908

```
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/048,964
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/057,650
;  EARLIER FILING DATE: 1997-09-05
;  EARLIER APPLICATION NUMBER: 60/056,884
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/057,669
;  EARLIER FILING DATE: 1997-09-05
;  EARLIER APPLICATION NUMBER: 60/049,610
;  EARLIER FILING DATE: 1997-06-13
;  EARLIER APPLICATION NUMBER: 60/061,060
;  EARLIER FILING DATE: 1997-10-02

  Query Match             0.2%;  Score 18;  DB 4;  Length 1541;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3192 TCTTCTTCATCCACATCA 3209
              ||||||||||||||||||
Db        384 TCTTCTTCATCCACATCA 367


RESULT 56
US-09-818-512-1/c
; Sequence 1, Application US/09818512
; Patent No. 6537780
; GENERAL INFORMATION:
;  APPLICANT: BEASLEY, Ellen et al.
;  TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001192
;  CURRENT APPLICATION NUMBER: US/09/818,512
;  CURRENT FILING DATE:  2001-03-28
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1781
;   TYPE: DNA
;   ORGANISM: Human
US-09-818-512-1

  Query Match             0.2%;  Score 18;  DB 4;  Length 1781;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3922 TATCTCTTTAAAAATATT 3939
              ||||||||||||||||||
Db       1409 TATCTCTTTAAAAATATT 1392


RESULT 57
US-08-203-905B-1
; Sequence 1, Application US/08203905B
; Patent No. 5646249
;  GENERAL INFORMATION:
;    APPLICANT:  KAYE, FEDERIC J.
;    APPLICANT:  OTTERSON, GREGORY A.
;    TITLE OF INVENTION:  ISOLATION AND CHARACTERIZATION OF A
;    TITLE OF INVENTION:  NOVEL CHAPERONE PROTEIN
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KNOBBE, MARTENS, OLSON & BEAR
;      STREET:  620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
;      CITY:  NEWPORT BEACH
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/203,905B
;      FILING DATE:  February 28, 1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KIRKPATRICK, ANITA M.
;      REGISTRATION NUMBER:  32,617
;      REFERENCE/DOCKET NUMBER:  NIH089.001A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-235-8550
;      TELEFAX:  619-235-0176
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2245 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  37..1452
US-08-203-905B-1

  Query Match             0.2%;  Score 18;  DB 1;  Length 2245;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4445 AATTCAACAATTGAAGCT 4462
              ||||||||||||||||||
Db        586 AATTCAACAATTGAAGCT 603


RESULT 58
US-09-482-273-20
; Sequence 20, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 71 Human Secreted Proteins
;  FILE REFERENCE: PZ030P1
;  CURRENT APPLICATION NUMBER: US/09/482,273
;  CURRENT FILING DATE: 2000-01-13
;  EARLIER APPLICATION NUMBER: PCT/US99/15849
;  EARLIER FILING DATE: 1999-07-14
;  EARLIER APPLICATION NUMBER: 60/092,921
;  EARLIER FILING DATE: 1998-07-15
;  EARLIER APPLICATION NUMBER: 60/092,922
;  EARLIER FILING DATE: 1998-07-15
;  EARLIER APPLICATION NUMBER: 60/092,956
;  EARLIER FILING DATE: 1998-07-15
;  NUMBER OF SEQ ID NOS: 267
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 2525
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (5)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (10)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1354)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-482-273-20
```

```
  Query Match             0.2%;  Score 18;  DB 4;  Length 2525;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6764 TCCTCAACAAGAAAAAAT 6781
              ||||||||||||||||||
Db        413 TCCTCAACAAGAAAAAAT 430


RESULT 59
US-09-252-991A-2112/c
; Sequence 2112, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 2112
;   LENGTH: 2598
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2112

  Query Match             0.2%;  Score 18;  DB 4;  Length 2598;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3062 CGCGGGCGCGCTCGGTGG 3079
              ||||||||||||||||||
Db        806 CGCGGGCGCGCTCGGTGG 789


RESULT 60
US-08-815-809-2
; Sequence 2, Application US/08815809
; Patent No. 6004777
; GENERAL INFORMATION:
;  APPLICANT: TARTAGLIA, James
;  APPLICANT: GOEBEL, Scott J.
;  APPLICANT: COX, William I.
;  APPLICANT: GETTIG, Russell R.
;  APPLICANT: PINCUS, Steven E.
;  APPLICANT: PAOLETTI, Enzo
;  APPLICANT: JACOBS, Bertram L.
;  TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND METHODS OF
;  TITLE OF INVENTION: MAKING AND USES THEREOF
;  FILE REFERENCE: 454310-3010
;  CURRENT APPLICATION NUMBER: US/08/815,809
;  CURRENT FILING DATE: 1997-03-12
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 2844
;   TYPE: DNA
;   ORGANISM: Vaccinia virus
US-08-815-809-2

  Query Match             0.2%;  Score 18;  DB 3;  Length 2844;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8680 CAACACTCCAGAGTCCCG 8697
              ||||||||||||||||||
Db        753 CAACACTCCAGAGTCCCG 770


RESULT 61
US-08-816-155B-4
; Sequence 4, Application US/08816155B
; Patent No. 5990091
; GENERAL INFORMATION:
;   APPLICANT:  TARTAGLIA, JAMES
;   APPLICANT:  COX, WILLIAM I.
;   APPLICANT:  GETTIG, RUSSELL R.
;   APPLICANT:  MARTINEZ, HECTOR
;   APPLICANT:  PAOLETTI, ENZO
;   APPLICANT:  PINCUS, STEVEN E.
;   TITLE OF INVENTION:  VECTORS HAVING ENHANCED EXPRESSION, AND
;   TITLE OF INVENTION:  METHODS OF MAKING AND USES THEREOF
;   NUMBER OF SEQUENCES:  48
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  FROMMER LAWRENCE & HAUG LLP
;     STREET:  745 FIFTH AVENUE
;     CITY:  NEW YORK
;     STATE:  NEW YORK
;     COUNTRY:  USA
;     ZIP:  10151
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/816,155B
;     FILING DATE:  12-MAR-1997
;     CLASSIFICATION:  514
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  KOWALSKI, THOMAS J.
;     REGISTRATION NUMBER:  32,147
;     REFERENCE/DOCKET NUMBER:  454310-2990
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  212-588-0800
;     TELEFAX:  212-588-0500
;  INFORMATION FOR SEQ ID NO:  4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  2856 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  DNA (genomic)
US-08-816-155B-4

  Query Match             0.2%;  Score 18;  DB 2;  Length 2856;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8680 CAACACTCCAGAGTCCCG 8697
              ||||||||||||||||||
Db        765 CAACACTCCAGAGTCCCG 782


RESULT 62
US-09-079-587-4
; Sequence 4, Application US/09079587
; Patent No. 6130066
; GENERAL INFORMATION:
;   APPLICANT:  TARTAGLIA, JAMES
;   APPLICANT:  COX, WILLIAM I.
;   APPLICANT:  GETTIG, RUSSELL R.
;   APPLICANT:  MARTINEZ, HECTOR
;   APPLICANT:  PAOLETTI, ENZO
;   APPLICANT:  PINCUS, STEVEN E.
;   TITLE OF INVENTION:  VECTORS HAVING ENHANCED EXPRESSION, AND
;   TITLE OF INVENTION:  METHODS OF MAKING AND USES THEREOF
;   NUMBER OF SEQUENCES:  48
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  FROMMER LAWRENCE & HAUG LLP
;     STREET:  745 FIFTH AVENUE
```

```
;       CITY:  NEW YORK
;       STATE:  NEW YORK
;       COUNTRY:  USA
;       ZIP:  10151
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/079,587
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/816,155
;       FILING DATE:  12-MAR-1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KOWALSKI, THOMAS J.
;       REGISTRATION NUMBER:  32,147
;       REFERENCE/DOCKET NUMBER:  454310-2990
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-588-0800
;       TELEFAX:  212-588-0500
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2856 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-09-079-587-4

  Query Match             0.2%;  Score 18;  DB 3;  Length 2856;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8680 CAACACTCCAGAGTCCCG 8697
             ||||||||||||||||||
Db        765 CAACACTCCAGAGTCCCG 782


RESULT 63
US-09-724-864-30/c
; Sequence 30, Application US/09724864
; Patent No. 6380362
; GENERAL INFORMATION:
;  APPLICANT:  Watson, James D
;  APPLICANT:  Murison, James G.
;  TITLE OF INVENTION:  Polynucleotides, polypeptides expressed
;  TITLE OF INVENTION:  by the polynucleotides and methods for their use.
;  FILE REFERENCE:  11000.1050U1
;  CURRENT APPLICATION NUMBER: US/09/724,864
;  CURRENT FILING DATE:  2000-11-28
;  PRIOR APPLICATION NUMBER:  U.S. No. 6380362 60/171,678
;  PRIOR FILING DATE:  1999-12-23
;  NUMBER OF SEQ ID NOS:  72
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 30
;   LENGTH: 2866
;   TYPE: DNA
;   ORGANISM: Mouse
US-09-724-864-30

  Query Match             0.2%;  Score 18;  DB 4;  Length 2866;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3412 TCGGCAGCATCCTGGGAG 3429
              ||||||||||||||||||
Db       1439 TCGGCAGCATCCTGGGAG 1422


RESULT 64
US-09-177-419C-3/c
; Sequence 3, Application US/09177419C
; Patent No. 6562609
; GENERAL INFORMATION:
;  APPLICANT: Russel, David W
;  APPLICANT:  Lund, Erik G
;  TITLE OF INVENTION: Cholesterol 25-Hydroxylase
;  FILE REFERENCE: UTSD1370
;  CURRENT APPLICATION NUMBER: US/09/177,419C
;  CURRENT FILING DATE:  1998-10-22
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 2951
;   TYPE: DNA
;   ORGANISM: Mouse
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1183)..(2076)
;   FEATURE:
;   OTHER INFORMATION: n signifies a, t, c or g.
US-09-177-419C-3

  Query Match             0.2%;  Score 18;  DB 4;  Length 2951;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5229 ATGGTGAGTGGGTGACAC 5246
              ||||||||||||||||||
Db       1781 ATGGTGAGTGGGTGACAC 1764


RESULT 65
US-08-021-601-1
; Sequence 1, Application US/08021601
; Patent No. 5591631
;  GENERAL INFORMATION:
;    APPLICANT:  Leppla, Stephen H.
;    APPLICANT:  Klimpel, Kurt R.
;    APPLICANT:  Nichols, Peter J.
;    APPLICANT:  Arora, Naveen
;    APPLICANT:  Singh, Yogendra
;    TITLE OF INVENTION:  ANTHRAX TOXIN FUSION PROTEINS AND
;    TITLE OF INVENTION:  RELATED METHODS
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Needle & Rosenberg, P.C.
;      STREET:  133 Carnegie Way, Suite 400
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/021,601
;      FILING DATE:  19930212
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spratt, Gwendolyn D.
;      REGISTRATION NUMBER:  36,016
;      REFERENCE/DOCKET NUMBER:  1414.057
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  404/688-0770
;      TELEFAX:  404/688-9880
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3291 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
```

```
;    TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Bacillus anthracis
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  580..2907
US-08-021-601-1

  Query Match             0.2%;  Score 18;  DB 1;  Length 3291;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7458 AAGCATACGTCAAAAAAC 7475
              ||||||||||||||||||
Db       3260 AAGCATACGTCAAAAAAC 3277


RESULT 66
US-08-082-849B-1
; Sequence 1, Application US/08082849B
; Patent No. 5677274
;  GENERAL INFORMATION:
;    APPLICANT:  Leppla, Stephen H.
;    APPLICANT:  Klimpel, Kurt R.
;    APPLICANT:  Arora, Naveen
;    APPLICANT:  Singh, Yogendra
;    APPLICANT:  Nichols, Peter J.
;    TITLE OF INVENTION:  Anthrax Toxin Fusion Proteins and
;    TITLE OF INVENTION:  Related Methods
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/082,849B
;      FILING DATE:  25-JUN-1993
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/021,601
;      FILING DATE:  12-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weber, Kenneth A.
;      REGISTRATION NUMBER:  31,677
;      REFERENCE/DOCKET NUMBER:  15280-161-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3291 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Bacillus anthracis
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  580..2907
;      OTHER INFORMATION:  /product= "Lethal Factor"
US-08-082-849B-1

  Query Match             0.2%;  Score 18;  DB 1;  Length 3291;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7458 AAGCATACGTCAAAAAAC 7475
              ||||||||||||||||||
Db       3260 AAGCATACGTCAAAAAAC 3277


RESULT 67
PCT-US94-01624-1
; Sequence 1, Application PC/TUS9401624
;  GENERAL INFORMATION:
;    APPLICANT:  Leppla, Stephen H.
;    APPLICANT:  Klimpel, Kurt R.
;    APPLICANT:  Arora, Naveen
;    APPLICANT:  Singh, Yogendra
;    APPLICANT:  Nichols, Peter J.
;    TITLE OF INVENTION:  ANTHRAX TOXIN FUSION PROTEINS AND
;    TITLE OF INVENTION:  RELATED METHODS
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND KHOURIE and CREW
;      STREET:  Steuart Street Tower, 20th Floor, One Market
;      STREET:  Plaza
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/01624
;      FILING DATE:  June 25, 1993
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weber, Kenneth A.
;      REGISTRATION NUMBER:  31,677
;      REFERENCE/DOCKET NUMBER:  15280-115
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 543-9600
;      TELEFAX:  (415) 543-5043
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3291 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Bacillus anthracis
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  580..2907
;      OTHER INFORMATION:  /product= "Lethal Factor"
PCT-US94-01624-1

  Query Match             0.2%;  Score 18;  DB 5;  Length 3291;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7458 AAGCATACGTCAAAAAAC 7475
              ||||||||||||||||||
Db       3260 AAGCATACGTCAAAAAAC 3277


RESULT 68
```

```
US-09-134-001C-1983
; Sequence 1983, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1983
;   LENGTH: 3636
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1983

  Query Match             0.2%;  Score 18;  DB 4;  Length 3636;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        426 AGAGCAAATTCAGCAACA 443
              ||||||||||||||||||
Db       2682 AGAGCAAATTCAGCAACA 2699


RESULT 69
US-08-464-961-1/c
; Sequence 1, Application US/08464961
; Patent No. 5824543
;  GENERAL INFORMATION:
;    APPLICANT:  Tung-Tien Sun
;    TITLE OF INVENTION:  Method for Expression and Isolation
;    TITLE OF INVENTION:  of        Biologically Active Molecules in Urine
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jane Massey Licata, Esq.
;      STREET:  210 Lake Drive East, Suite 201
;      CITY:  Cherry Hill
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08002
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;      COMPUTER:  IBM 486
;      OPERATING SYSTEM:  WINDOWS FOR WORKGROUPS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/464,961
;      FILING DATE:  Herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jane Massey Licata
;      REGISTRATION NUMBER:  32,257
;      REFERENCE/DOCKET NUMBER:  NYU-0003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (609) 779-2400
;      TELEFAX:  (609) 779-8488
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3963
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    ANTI-SENSE:  No
US-08-464-961-1

  Query Match             0.2%;  Score 18;  DB 1;  Length 3963;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2376 AAATAAAAACAAGCAAAT 2393
              ||||||||||||||||||
Db       3817 AAATAAAAACAAGCAAAT 3800


RESULT 70
US-08-907-800A-1/c
; Sequence 1, Application US/08907800A
; Patent No. 6001646
;  GENERAL INFORMATION:
;    APPLICANT:  SUN, Tung-Tien
;    TITLE OF INVENTION:  METHOD AND VECTOR FOR EXPRESSION AND
;    TITLE OF INVENTION:  ISOLATION OF BIOLOGICALLY ACTIVE MOLECULES IN URINE
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK, P.L.L.C.
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/907,800A
;      FILING DATE:  08-AUG-1997
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/464,961
;      FILING DATE:  05-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROWDY, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  SUN=9A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3963 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-907-800A-1

  Query Match             0.2%;  Score 18;  DB 3;  Length 3963;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2376 AAATAAAAACAAGCAAAT 2393
              ||||||||||||||||||
Db       3817 AAATAAAAACAAGCAAAT 3800


RESULT 71
US-09-083-541-1/c
; Sequence 1, Application US/09083541
; Patent No. 6323390
;  GENERAL INFORMATION:
;    APPLICANT:  WU, Xue-Ru
;    APPLICANT:  SUN, Tung-Tien
;    TITLE OF INVENTION:  TRANSGENIC ANIMAL MODELS FOR HUMAN
;    TITLE OF INVENTION:  BLADDER CANCER
;    NUMBER OF SEQUENCES:  8
```

```
;  CORRESPONDENCE ADDRESS:
;    ADDRESSEE:  BROWDY AND NEIMARK, P.L.L.C.
;    STREET:  419 Seventh Street, N.W., Suite 300
;    CITY:  Washington
;    STATE:  D.C.
;    COUNTRY:  USA
;    ZIP:  20004
;  COMPUTER READABLE FORM:
;    MEDIUM TYPE:  Floppy disk
;    COMPUTER:  IBM PC compatible
;    OPERATING SYSTEM:  PC-DOS/MS-DOS
;    SOFTWARE:  PatentIn Release #1.0, Version #1.30
;  CURRENT APPLICATION DATA:
;    APPLICATION NUMBER:  US/09/083,541
;    FILING DATE:
;    CLASSIFICATION:
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER:  US 08/969,315
;    FILING DATE:  13-NOV-1997
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER:  US 08/907,800
;    FILING DATE:  08-AUG-1997
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER:  US 08/464,961
;    FILING DATE:  05-JUN-1995
;  ATTORNEY/AGENT INFORMATION:
;    NAME:  YUN, Allen C.
;    REGISTRATION NUMBER:  37,971
;    REFERENCE/DOCKET NUMBER:  WU=42
;  TELECOMMUNICATION INFORMATION:
;    TELEPHONE:  202-628-5197
;    TELEFAX:  202-737-3528
; INFORMATION FOR SEQ ID NO:  1:
;  SEQUENCE CHARACTERISTICS:
;    LENGTH:  3963 base pairs
;    TYPE:  nucleic acid
;    STRANDEDNESS:  single
;    TOPOLOGY:  linear
;  MOLECULE TYPE:  cDNA
US-09-083-541-1

  Query Match             0.2%;  Score 18;  DB 4;  Length 3963;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2376 AAATAAAAACAAGCAAAT 2393
              ||||||||||||||||||
Db       3817 AAATAAAAACAAGCAAAT 3800


RESULT 72
US-08-969-315-1/c
; Sequence 1, Application US/08969315
; Patent No. 6339183
;  GENERAL INFORMATION:
;    APPLICANT:  Tung-Tien Sun
;    TITLE OF INVENTION:  Method and Vector for Expression of Genes
;    TITLE OF INVENTION:  in Bladder Epithelium and Isolation of Biologically
;    TITLE OF INVENTION:  Active Molecules in Urine
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jane Massey Licata, Esq.
;      STREET:  66 E. Main Street
;      CITY:  Marlton
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08053
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;      COMPUTER:  IBM 486
;      OPERATING SYSTEM:  WINDOWS FOR WORKGROUPS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/969,315
;      FILING DATE:  Herewith
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/464,961
;      FILING DATE:  June 5, 1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/907,800
;      FILING DATE:  August 8, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jane Massey Licata
;      REGISTRATION NUMBER:  32,257
;      REFERENCE/DOCKET NUMBER:  NYU-0032
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (609) 779-2400
;      TELEFAX:  (609) 810-1454
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3963
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    ANTI-SENSE:  No
US-08-969-315-1

  Query Match             0.2%;  Score 18;  DB 4;  Length 3963;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2376 AAATAAAAACAAGCAAAT 2393
              ||||||||||||||||||
Db       3817 AAATAAAAACAAGCAAAT 3800


RESULT 73
PCT-US96-08233-1/c
; Sequence 1, Application PC/TUS9608233
;  GENERAL INFORMATION:
;    APPLICANT:  Tung-Tien Sun
;    TITLE OF INVENTION:  Method for Expression and
;    TITLE OF INVENTION:  and Isolation of Biologically Active Molecules
;    TITLE OF INVENTION:  in Urine
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jane Massey Licata, Esq.
;      STREET:  210 Lake Drive East, Suite 201
;      CITY:  Cherry Hill
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08002
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;      COMPUTER:  IBM 486
;      OPERATING SYSTEM:  WINDOWS FOR WORKGROUPS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/08233
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jane Massey Licata
;      REGISTRATION NUMBER:  32,257
;      REFERENCE/DOCKET NUMBER:  NYU-0005
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (609) 779-2400
;      TELEFAX:  (609) 779-8488
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3963
;      TYPE:  Nucleic Acid
```

```
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    ANTI-SENSE:  No
PCT-US96-08233-1

  Query Match             0.2%;  Score 18;  DB 5;  Length 3963;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2376 AAATAAAAACAAGCAAAT 2393
              ||||||||||||||||||
Db       3817 AAATAAAAACAAGCAAAT 3800


RESULT 74
US-08-107-748-3
; Sequence 3, Application US/08107748
; Patent No. 5576428
;  GENERAL INFORMATION:
;    APPLICANT:  Fitzmaurice, Leona
;    APPLICANT:  Mirkov, Erik
;    APPLICANT:  Butler, William O.
;    APPLICANT:  Dickinson, Craig D.
;    APPLICANT:  Elliott, Kathryn J.
;    APPLICANT:  Konno, Yoshihiro
;    APPLICANT:  Holtz, Gregory C.
;    TITLE OF INVENTION:  NOVEL INVERTASE GENE(S) AND USES THEREOF
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/107,748
;      FILING DATE:  20-AUG-1993
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/660,344
;      FILING DATE:  22-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/771,331
;      FILING DATE:  04-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L.
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  P51 9503
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-546-4737
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4032 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Lycopersicon esculentum
;    FEATURE:
;      NAME/KEY:  precursor_RNA
;      LOCATION:  889
US-08-107-748-3

  Query Match             0.2%;  Score 18;  DB 1;  Length 4032;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5324 TATTTATTTTTATAACAT 5341
              ||||||||||||||||||
Db       1025 TATTTATTTTTATAACAT 1042


RESULT 75
US-08-245-809-4
; Sequence 4, Application US/08245809
; Patent No. 5665579
;  GENERAL INFORMATION:
;    APPLICANT:  Fitzmaurice, Leona
;    APPLICANT:  Mirkov, Erik
;    APPLICANT:  Butler, William O.
;    APPLICANT:  Dickinson, Craig D.
;    APPLICANT:  Elliott, Kathryn J.
;    APPLICANT:  Konno, Yoshihiro
;    TITLE OF INVENTION:  NOVEL INVERTASE GENE(S) AND USES THEREOF
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fitch, Even, Tabin & Flannery
;      STREET:  135 South LaSalle Street, Suite 900
;      CITY:  Chicago
;      STATE:  IL
;      COUNTRY:  U.S.A.
;      ZIP:  60603-4277
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/245,809
;      FILING DATE:  17-MAY-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/771,331
;      FILING DATE:  03-OCT-1991
;      APPLICATION NUMBER:  US 07/660,344
;      FILING DATE:  22-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L.
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  51514
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-552-1311
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4032 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Lycopersicon esculentum
;    FEATURE:
;      NAME/KEY:  prim_transcript
;      LOCATION:  889
US-08-245-809-4

  Query Match             0.2%;  Score 18;  DB 1;  Length 4032;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5324 TATTTATTTTTATAACAT 5341
              ||||||||||||||||||
Db       1025 TATTTATTTTTATAACAT 1042


RESULT 76
PCT-US92-01385-3
; Sequence 3, Application PC/TUS9201385
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Fitzmaurice, Leona
;    APPLICANT:  Mirkov, Erik
;    APPLICANT:  Butler, William O.
;    APPLICANT:  Dickinson, Craig D.
;    APPLICANT:  Elliott, Kathryn J.
;    APPLICANT:  Konno, Yoshihiro
;    APPLICANT:  Holtz, Gregory C.
;    TITLE OF INVENTION:  NOVEL INVERTASE GENE(S) AND USES THEREOF
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fitch, Even, Tabin & Flannery
;      STREET:  135 South LaSalle Street, Suite 900
;      CITY:  Chicago
;      STATE:  IL
;      COUNTRY:  U.S.A.
;      ZIP:  60603-4277
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/01385
;      FILING DATE:  19920221
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/660,344
;      FILING DATE:  22-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/771,331
;      FILING DATE:  04-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L.
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  52498PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-552-1311
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4032 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Lycopersicon esculentum
;    FEATURE:
;      NAME/KEY:  precursor_RNA
;      LOCATION:  889
PCT-US92-01385-3

  Query Match             0.2%;  Score 18;  DB 5;  Length 4032;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5324 TATTTATTTTTATAACAT 5341
              ||||||||||||||||||
Db       1025 TATTTATTTTTATAACAT 1042


RESULT 77
US-08-201-697-3/c
; Sequence 3, Application US/08201697
; Patent No. 5705623
;  GENERAL INFORMATION:
;    APPLICANT:  Wiggins, Roger C.
;    APPLICANT:  Thomas, Peedikayil E.
;    TITLE OF INVENTION:  Mammalian Glomerular Epithelial Protein
;    TITLE OF INVENTION:  1
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/201,697
;      FILING DATE:  25-FEB-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryb A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-UM 9783
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4752 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  808..4372
US-08-201-697-3

  Query Match             0.2%;  Score 18;  DB 1;  Length 4752;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3192 TCTTCTTCATCCACATCA 3209
              ||||||||||||||||||
Db       4321 TCTTCTTCATCCACATCA 4304


RESULT 78
US-08-201-697-6/c
; Sequence 6, Application US/08201697
; Patent No. 5705623
;  GENERAL INFORMATION:
;    APPLICANT:  Wiggins, Roger C.
;    APPLICANT:  Thomas, Peedikayil E.
;    TITLE OF INVENTION:  Mammalian Glomerular Epithelial Protein
;    TITLE OF INVENTION:  1
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/201,697
;      FILING DATE:  25-FEB-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryb A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-UM 9783
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
```

```
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4799 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-201-697-6

  Query Match             0.2%;  Score 18;  DB 1;  Length 4799;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3192 TCTTCTTCATCCACATCA 3209
              ||||||||||||||||||
Db       4368 TCTTCTTCATCCACATCA 4351


RESULT 79
US-08-201-697-5/c
; Sequence 5, Application US/08201697
; Patent No. 5705623
;  GENERAL INFORMATION:
;    APPLICANT:  Wiggins, Roger C.
;    APPLICANT:  Thomas, Peedikayil E.
;    TITLE OF INVENTION:  Mammalian Glomerular Epithelial Protein
;    TITLE OF INVENTION:  1
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/201,697
;      FILING DATE:  25-FEB-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryb A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-UM 9783
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4815 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-201-697-5

  Query Match             0.2%;  Score 18;  DB 1;  Length 4815;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3192 TCTTCTTCATCCACATCA 3209
              ||||||||||||||||||
Db       4384 TCTTCTTCATCCACATCA 4367


RESULT 80
US-08-201-697-1/c
; Sequence 1, Application US/08201697
; Patent No. 5705623
;  GENERAL INFORMATION:
;    APPLICANT:  Wiggins, Roger C.
;    APPLICANT:  Thomas, Peedikayil E.
;    TITLE OF INVENTION:  Mammalian Glomerular Epithelial Protein
;    TITLE OF INVENTION:  1
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/201,697
;      FILING DATE:  25-FEB-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryb A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-UM 9783
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5679 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  180..3740
US-08-201-697-1

  Query Match             0.2%;  Score 18;  DB 1;  Length 5679;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3192 TCTTCTTCATCCACATCA 3209
              ||||||||||||||||||
Db       3690 TCTTCTTCATCCACATCA 3673


RESULT 81
US-08-815-809-3/c
; Sequence 3, Application US/08815809
; Patent No. 6004777
; GENERAL INFORMATION:
;  APPLICANT: TARTAGLIA, James
;  APPLICANT: GOEBEL, Scott J.
;  APPLICANT: COX, William I.
;  APPLICANT: GETTIG, Russell R.
;  APPLICANT: PINCUS, Steven E.
;  APPLICANT: PAOLETTI, Enzo
;  APPLICANT: JACOBS, Bertram L.
;  TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND METHODS OF
;  TITLE OF INVENTION: MAKING AND USES THEREOF
;  FILE REFERENCE: 454310-3010
;  CURRENT APPLICATION NUMBER: US/08/815,809
;  CURRENT FILING DATE: 1997-03-12
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 6628
;   TYPE: DNA
;   ORGANISM: Vaccinia virus
US-08-815-809-3

  Query Match             0.2%;  Score 18;  DB 3;  Length 6628;
```

```
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8680 CAACACTCCAGAGTCCCG 8697
             ||||||||||||||||||
Db       3082 CAACACTCCAGAGTCCCG 3065


RESULT 82
US-08-816-155B-5/c
; Sequence 5, Application US/08816155B
; Patent No. 5990091
;  GENERAL INFORMATION:
;    APPLICANT:  TARTAGLIA, JAMES
;    APPLICANT:  COX, WILLIAM I.
;    APPLICANT:  GETTIG, RUSSELL R.
;    APPLICANT:  MARTINEZ, HECTOR
;    APPLICANT:  PAOLETTI, ENZO
;    APPLICANT:  PINCUS, STEVEN E.
;    TITLE OF INVENTION:  VECTORS HAVING ENHANCED EXPRESSION, AND
;    TITLE OF INVENTION:  METHODS OF MAKING AND USES THEREOF
;    NUMBER OF SEQUENCES:  48
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FROMMER LAWRENCE & HAUG LLP
;      STREET:  745 FIFTH AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10151
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/816,155B
;      FILING DATE:  12-MAR-1997
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOWALSKI, THOMAS J.
;      REGISTRATION NUMBER:  32,147
;      REFERENCE/DOCKET NUMBER:  454310-2990
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-588-0800
;      TELEFAX:  212-588-0500
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6649 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-816-155B-5

  Query Match             0.2%;  Score 18;  DB 2;  Length 6649;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8680 CAACACTCCAGAGTCCCG 8697
             ||||||||||||||||||
Db       3103 CAACACTCCAGAGTCCCG 3086


RESULT 83
US-09-079-587-5/c
; Sequence 5, Application US/09079587
; Patent No. 6130066
;  GENERAL INFORMATION:
;    APPLICANT:  TARTAGLIA, JAMES
;    APPLICANT:  COX, WILLIAM I.
;    APPLICANT:  GETTIG, RUSSELL R.
;    APPLICANT:  MARTINEZ, HECTOR
;    APPLICANT:  PAOLETTI, ENZO
;    APPLICANT:  PINCUS, STEVEN E.
;    TITLE OF INVENTION:  VECTORS HAVING ENHANCED EXPRESSION, AND
;    TITLE OF INVENTION:  METHODS OF MAKING AND USES THEREOF
;    NUMBER OF SEQUENCES:  48
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FROMMER LAWRENCE & HAUG LLP
;      STREET:  745 FIFTH AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10151
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/079,587
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/816,155
;      FILING DATE:  12-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOWALSKI, THOMAS J.
;      REGISTRATION NUMBER:  32,147
;      REFERENCE/DOCKET NUMBER:  454310-2990
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-588-0800
;      TELEFAX:  212-588-0500
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6649 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-079-587-5

  Query Match             0.2%;  Score 18;  DB 3;  Length 6649;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8680 CAACACTCCAGAGTCCCG 8697
             ||||||||||||||||||
Db       3103 CAACACTCCAGAGTCCCG 3086


RESULT 84
US-08-135-511-31
; Sequence 31, Application US/08135511
; Patent No. 5558999
;  GENERAL INFORMATION:
;    APPLICANT:  Chiang, John
;    TITLE OF INVENTION:  Cholesterol 7a-Hydroxylase Gene
;    TITLE OF INVENTION:  Regulatory Elements and Methods for Using Them
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/135,511
;      FILING DATE:  13-OCT-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
```

```
;       NAME:  SANDERCOCK, Colin G.
;       REGISTRATION NUMBER:  31,298
;       REFERENCE/DOCKET NUMBER:  18748/175
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202)672-5300
;       TELEFAX:  (202)672-5399
;       TELEX:  904136
;   INFORMATION FOR SEQ ID NO:  31:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  7970 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-135-511-31

  Query Match             0.2%;  Score 18;  DB 1;  Length 7970;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2380 AAAAACAAGCAAATGGAA 2397
             ||||||||||||||||||
Db       4974 AAAAACAAGCAAATGGAA 4991


RESULT 85
US-08-187-453-31
; Sequence 31, Application US/08187453
; Patent No. 5753431
;  GENERAL INFORMATION:
;     APPLICANT:  Chiang, John
;     TITLE OF INVENTION:  Cholesterol 7a-Hydroxylase Gene
;     TITLE OF INVENTION:  Regulatory Elements and Transcription Factors
;     NUMBER OF SEQUENCES:  37
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Foley & Lardner
;       STREET:  3000 K Street, N.W., Suite 500
;       CITY:  Washington, D.C.
;       COUNTRY:  USA
;       ZIP:  20007-5109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/187,453
;       FILING DATE:  28-JAN-1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/135,488
;       FILING DATE:  13-OCT-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/135,511
;       FILING DATE:  13-OCT-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/135,510
;       FILING DATE:  13-OCT-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  SANDERCOCK, Colin G.
;       REGISTRATION NUMBER:  31,298
;       REFERENCE/DOCKET NUMBER:  18748/188
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202)672-5300
;       TELEFAX:  (202)672-5399
;       TELEX:  904136
;   INFORMATION FOR SEQ ID NO:  31:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  7997 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-187-453-31

  Query Match             0.2%;  Score 18;  DB 1;  Length 7997;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2380 AAAAACAAGCAAATGGAA 2397
             ||||||||||||||||||
Db       5001 AAAAACAAGCAAATGGAA 5018


RESULT 86
US-08-562-985A-7
; Sequence 7, Application US/08562985A
; Patent No. 5821057
;  GENERAL INFORMATION:
;     APPLICANT:  CHIANG, John Y.L.
;     APPLICANT:  STROUP, Diane
;     TITLE OF INVENTION:  AN ASSAY FOR AGENTS THAT AFFECT
;     TITLE OF INVENTION:  CHOLESTEROL 7ALPHA-HYDROXYLASE EXPRESSION AND A
;     TITLE OF INVENTION:  CHARACTERIZATION OF ITS REGULATORY ELEMENTS
;     NUMBER OF SEQUENCES:  8
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Foley & Lardner
;       STREET:  3000 K Street, N.W., Suite 500
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  USA
;       ZIP:  20007-5109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/562,985A
;       FILING DATE:  27-NOV-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  SANDERCOCK, Colin G.
;       REGISTRATION NUMBER:  31,298
;       REFERENCE/DOCKET NUMBER:  18748/213/HOCE
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202)672-5300
;       TELEFAX:  (202)672-5399
;       TELEX:  904136
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  7997 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-562-985A-7

  Query Match             0.2%;  Score 18;  DB 1;  Length 7997;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2380 AAAAACAAGCAAATGGAA 2397
             ||||||||||||||||||
Db       5001 AAAAACAAGCAAATGGAA 5018


RESULT 87
US-08-961-527-66
; Sequence 66, Application US/08961527
; Patent No. 6420135
;  GENERAL INFORMATION:
;     APPLICANT:  Charles Kunsch
;     TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;     NUMBER OF SEQUENCES:  391
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Human Genome Sciences, Inc.
```

```
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  66:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10726 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-66

  Query Match             0.2%;  Score 18;  DB 4;  Length 10726;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2805 CCTGCAGTTGATCCAGCA 2822
              ||||||||||||||||||
Db       7352 CCTGCAGTTGATCCAGCA 7369


RESULT 88
US-08-961-527-33
; Sequence 33, Application US/08961527
; Patent No. 6420135
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  13206 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-33

  Query Match             0.2%;  Score 18;  DB 4;  Length 13206;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5326 TTTATTTTTATAACATCA 5343
              ||||||||||||||||||
Db       6453 TTTATTTTTATAACATCA 6470


RESULT 89
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
```

```
;      LOCATION:  36155..41830
US-08-804-227C-7

  Query Match             0.2%;  Score 18;  DB 2;  Length 44377;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8701 GGCGGCCGGCTCTCCGCC 8718
              ||||||||||||||||||
Db      33543 GGCGGCCGGCTCTCCGCC 33560


RESULT 90
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-198-1

  Query Match             0.2%;  Score 18;  DB 2;  Length 44377;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8701 GGCGGCCGGCTCTCCGCC 8718
              ||||||||||||||||||
Db      33543 GGCGGCCGGCTCTCCGCC 33560


RESULT 91
US-08-764-233A-1
; Sequence 1, Application US/08764233A
; Patent No. 5716849
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Beck, James J.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Neff, Snezanna
;    APPLICANT:  Ryals, John A.
;    TITLE OF INVENTION:  Genes For The Biosynthesis Of Soraphen
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,233A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/729,214
;      FILING DATE:  09-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,261
;      FILING DATE:  08-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  1506/CIP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 541-8587
;      TELEFAX:  (919) 541-8689
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  49377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Sorangium cellulosum
;    IMMEDIATE SOURCE:
;      CLONE:  p98/1, pJL3, and pVKM15
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  383..760
;      OTHER INFORMATION:  /product= "SorR"
;      OTHER INFORMATION:  /note= "This gene encodes a protein that is highly homolog
;      OTHER INFORMATION:  the reductase domains of type I PKSs such as eryA from
;      OTHER INFORMATION:  Saccharopolyspora erythraea."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  927..19874
;      OTHER INFORMATION:  /product= "SorA"
;      OTHER INFORMATION:  /note= "Gene product is highly homologous to type I PKSs t
;      OTHER INFORMATION:  are known to be involved in the synthesis of polyketide
;      OTHER INFORMATION:  compounds."
```

```
;   FEATURE:
;     NAME/KEY:  misc_feature
;     LOCATION:  942..7115
;     OTHER INFORMATION:  /product= "Module 1 of SorA"
;   FEATURE:
;     NAME/KEY:  misc_feature
;     LOCATION:  7203..12884
;     OTHER INFORMATION:  /product= "Module 2 of SorA"
;   FEATURE:
;     NAME/KEY:  misc_feature
;     LOCATION:  13455..19616
;     OTHER INFORMATION:  /product= "Module 3 of SorA"
;   FEATURE:
;     NAME/KEY:  misc_feature
;     LOCATION:  19871..46318
;     OTHER INFORMATION:  /product= "SorB"
;     OTHER INFORMATION:  /note= "Gene product is highly homologous to type I PKS genes
;   FEATURE:
;     NAME/KEY:  misc_feature
;     LOCATION:  19870..24556
;     OTHER INFORMATION:  /product= "Module 1 of SorB"
;   FEATURE:
;     NAME/KEY:  misc_feature
;     LOCATION:  24638..30820
;     OTHER INFORMATION:  /product= "Module 2 of SorB"
;   FEATURE:
;     NAME/KEY:  misc_feature
;     LOCATION:  30881..35446
;     OTHER INFORMATION:  /product= "Module 3 of SorB"
;   FEATURE:
;     NAME/KEY:  misc_feature
;     LOCATION:  35528..40114
;     OTHER INFORMATION:  /product= "Module 4 of SorB"
;   FEATURE:
;     NAME/KEY:  misc_feature
;     LOCATION:  40190..46318
;     OTHER INFORMATION:  /product= "Module 5 of SorB"
;   FEATURE:
;     NAME/KEY:  misc_feature
;     LOCATION:  46851..47891
;     OTHER INFORMATION:  /product= "SorM"
;     OTHER INFORMATION:  /note= "The protein encoded by the sorM gene is highly
;     OTHER INFORMATION:  homologous to the methyltransferase from Streptomyces
;     OTHER INFORMATION:  hygroscopicus that is involved in the synthesis of the
;     OTHER INFORMATION:  polyketide rappamicin."
US-08-764-233A-1

  Query Match             0.2%;  Score 18;  DB 1;  Length 49377;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5817 CCTGGCGGCGCTCGCGCA 5834
             ||||||||||||||||||
Db       6662 CCTGGCGGCGCTCGCGCA 6679


RESULT 92
US-09-245-041-5/c
; Sequence 5, Application US/09245041
; Patent No. 6274339
; GENERAL INFORMATION:
;  APPLICANT: Moore, K.
;  APPLICANT: Nagle, D.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT
;  TITLE OF INVENTION: OF BODY WEIGHT DISORDERS INCLUDING OBESITY
;  FILE REFERENCE: 7853-136
;  CURRENT APPLICATION NUMBER: US/09/245,041
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/093,630
;  EARLIER FILING DATE: 1998-07-21
;  EARLIER APPLICATION NUMBER: 60/104,978
;  EARLIER FILING DATE: 1998-10-20
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 90050
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-245-041-5

  Query Match             0.2%;  Score 18;  DB 3;  Length 90050;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        224 CTGGTGCTGGGAAAAGTA 241
              ||||||||||||||||||
Db      32338 CTGGTGCTGGGAAAAGTA 32321


RESULT 93
US-09-818-512-3/c
; Sequence 3, Application US/09818512
; Patent No. 6537780
; GENERAL INFORMATION:
;  APPLICANT: BEASLEY, Ellen et al.
;  TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001192
;  CURRENT APPLICATION NUMBER: US/09/818,512
;  CURRENT FILING DATE:  2001-03-28
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 116592
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(116592)
;   OTHER INFORMATION: n = A,T,C or G
US-09-818-512-3

  Query Match             0.2%;  Score 18;  DB 4;  Length 116592;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3922 TATCTCTTTAAAAATATT 3939
              ||||||||||||||||||
Db     114304 TATCTCTTTAAAAATATT 114287


RESULT 94
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
;  APPLICANT: Pan, Yang
;  TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
;  TITLE OF INVENTION: AND USES THEREOF
;  FILE REFERENCE: 09404/052001
;  CURRENT APPLICATION NUMBER: US/09/128,155
;  CURRENT FILING DATE: 1998-08-03
;  EARLIER APPLICATION NUMBER: US 60/091,650
;  EARLIER FILING DATE: 1998-07-02
;  EARLIER APPLICATION NUMBER: US 60/054,646
;  EARLIER FILING DATE: 1997-08-04
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 16
;   LENGTH: 152331
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(152331)
```

```
;   OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

  Query Match             0.2%;  Score 18;  DB 3;  Length 152331;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2382 AAACAAGCAAATGGAAAA 2399
              ||||||||||||||||||
Db      37698 AAACAAGCAAATGGAAAA 37715


RESULT 95
US-08-545-528D-1
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
;  APPLICANT: Fraser et al.
;  TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
; Patent No. 6537773
;  TITLE OF INVENTION:  Thereof, and Uses Thereof
;  FILE REFERENCE: PB193P1
;  CURRENT APPLICATION NUMBER: US/08/545,528D
;  CURRENT FILING DATE:  1995-10-19
;  PRIOR APPLICATION NUMBER: US 08/488,018
;  PRIOR FILING DATE: 1995-06-07
;  PRIOR APPLICATION NUMBER: US 08/473,545
;  PRIOR FILING DATE: 1995-06-07
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 580073
;   TYPE: DNA
;   ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

  Query Match             0.2%;  Score 18;  DB 4;  Length 580073;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4609 TTTGTTAAATGAAAAGAA 4626
              ||||||||||||||||||
Db      77875 TTTGTTAAATGAAAAGAA 77892


RESULT 96
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234220)..(234220)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234814)..(234814)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309398)..(309398)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309418)..(309418)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312837)..(312837)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312993)..(312993)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (319226)..(319226)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559167)..(559167)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559241)..(559241)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (600992)..(600992)
```

```
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (622708)..(622708)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657081)..(657081)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657203)..(657203)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (674435)..(674435)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (682442)..(682442)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (713652)..(713652)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (741684)..(741684)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779455)..(779455)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779676)..(779676)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (855539)..(855539)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (871619)..(871619)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1084830)..(1084830)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1096846)..(1096846)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1119881)..(1119881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1130881)..(1130881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1310988)..(1310988)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1313224)..(1313224)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349473)..(1349473)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349491)..(1349491)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1470091)..(1470091)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1569020)..(1569020)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1602912)..(1602912)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1603734)..(1603734)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1637998)..(1637998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1664854)..(1664855)
;   OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

  Query Match             0.2%;  Score 18;  DB 4;  Length 1664976;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5935 TGCATATATATCTGTTGG 5952
             ||||||||||||||||||
Db    1538766 TGCATATATATCTGTTGG 1538783


RESULT 97
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             0.2%;  Score 18;  DB 3;  Length 4403765;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8667 CGGCCGATATCGCCAACA 8684
             ||||||||||||||||||
Db     176989 CGGCCGATATCGCCAACA 176972


RESULT 98
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             0.2%;  Score 18;  DB 3;  Length 4411529;
```

```
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8667 CGGCCGATATCGCCAACA 8684
              ||||||||||||||||||
Db     176820 CGGCCGATATCGCCAACA 176803


RESULT 99
US-08-173-510B-57/c
; Sequence 57, Application US/08173510B
; Patent No. 5747296
;  GENERAL INFORMATION:
;    APPLICANT:  MATTHEW MOYLE, ET AL.
;    TITLE OF INVENTION:  NOVEL NEUTROPHIL INHIBITORS
;    NUMBER OF SEQUENCES:  104
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/173,510B
;      FILING DATE:  23-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/151,064
;      FILING DATE:  10-NOV-1993
;      APPLICATION NUMBER:  08/060,433
;      FILING DATE:  11-MAY-1993
;      APPLICATION NUMBER:  07/996,972
;      FILING DATE:  24-DEC-1992
;      APPLICATION NUMBER:  07/881,721
;      FILING DATE:  11-MAY-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BIGGS, SUZANNE L.
;      REGISTRATION NUMBER:  30,158
;      REFERENCE/DOCKET NUMBER:  205/073
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  57:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  38
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  NUCLEIC
US-08-173-510B-57

  Query Match             0.2%;  Score 17;  DB 1;  Length 38;
  Best Local Similarity   100.0%;  Pred. No. 3.5e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4919 ACCCGAAAATACAGAAA 4935
              |||||||||||||||||
Db         32 ACCCGAAAATACAGAAA 16


RESULT 100
US-08-458-218-57/c
; Sequence 57, Application US/08458218
; Patent No. 5789178
;  GENERAL INFORMATION:
;    APPLICANT:  MATTHEW MOYLE ET AL.
;    TITLE OF INVENTION:  NOVEL NEUTROPHIL INHIBITORS
;    NUMBER OF SEQUENCES:  102
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,218
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/151,064
;      FILING DATE:  10-NOVEMBER-1993
;      APPLICATION NUMBER:  08/060,433
;      FILING DATE:  11-MAY-1993
;      APPLICATION NUMBER:  07/881,721
;      FILING DATE:  11-MAY-1992
;      APPLICATION NUMBER:  07/996,972
;      FILING DATE:  24-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BIGGS, SUZANNE L.
;      REGISTRATION NUMBER:  30,158
;      REFERENCE/DOCKET NUMBER:  203/226
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  57:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  38
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  NUCLEIC
US-08-458-218-57

  Query Match             0.2%;  Score 17;  DB 1;  Length 38;
  Best Local Similarity   100.0%;  Pred. No. 3.5e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4919 ACCCGAAAATACAGAAA 4935
              |||||||||||||||||
Db         32 ACCCGAAAATACAGAAA 16


Search completed: August 22, 2003, 14:48:10
Job time : 570 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 21, 2003, 21:25:22 ; Search time 2054 Seconds
(without alignments)
11924.051 Million cell updates/sec

Title: US-09-831-000-1_COPY_60966_70038
Perfect score: 9073
Sequence: 1 atggagagttccgtcggatg.....gcggccacggatcgccttga 9073

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 300 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	9073	100.0	128139	24	AAI64291	RRV genome nucleot
	2	9073	100.0	133719	21	AAC64754	Macaca mulatta rha
c	3	22	0.2	12590	24	ABL33920	Human immune syste
c	4	21	0.2	3628	25	ABZ09971	Haematopoietic cel
c	5	21	0.2	3628	25	ABZ10117	Haematopoietic cel
c	6	21	0.2	6127	24	ABL32593	Human immune syste
c	7	21	0.2	36383	23	ABL05364	Drosophila melanog
	8	21	0.2	124884	22	AAH74201	Nucleotide sequenc
	9	21	0.2	125157	22	AAH74202	Nucleotide sequenc
c	10	20	0.2	255	24	ABK80864	Bacillus clausii g
c	11	20	0.2	310	23	ABV62066	Human prostate exp
	12	20	0.2	356	24	ABK79195	Bacillus clausii g
c	13	20	0.2	456	22	ABA14638	Human nervous syst
	14	20	0.2	461	24	ABN94325	Gene #823 used to
c	15	20	0.2	483	22	ABA14637	Human nervous syst
	16	20	0.2	736	22	AAI95250	Human neuroblastom
	17	20	0.2	774	24	ABK71638	Human dithp polynu
	18	20	0.2	834	24	ABL91226	Chlamydia pneumoni
c	19	20	0.2	848	24	ABQ29020	Oligonucleotide fo
	20	20	0.2	848	24	ABQ29021	Oligonucleotide fo
	21	20	0.2	945	24	ABN74144	Bovine embryonic g
c	22	20	0.2	960	21	AAZ45345	DNA encoding a Bac
c	23	20	0.2	963	24	ABK80177	Bacillus clausii g
c	24	20	0.2	1472	24	ABQ91912	Human NF-kB activa
c	25	20	0.2	1839	21	AAC61881	cDNA encoding a hu
c	26	20	0.2	1904	19	AAV62752	Human secreted pro
c	27	20	0.2	1904	24	ABQ92055	Human polynucleoti
c	28	20	0.2	1924	24	ABQ91913	Human NF-kB activa
c	29	20	0.2	1925	22	AAF30021	Human hydrophobic
c	30	20	0.2	1927	22	AAF93751	Human cDNA encodin
c	31	20	0.2	1977	21	AAA96358	cDNA encoding a no
c	32	20	0.2	1977	22	AAS46176	Human DNA encoding
c	33	20	0.2	1977	24	ABL95723	Human angiogenesis
c	34	20	0.2	1977	24	ABK69983	cDNA encoding huma
c	35	20	0.2	1977	24	ABL88234	Human PRO4499 cDNA
c	36	20	0.2	1977	25	ACA06124	cDNA encoding huma
c	37	20	0.2	1977	25	ACA06181	cDNA encoding huma
c	38	20	0.2	1977	25	ACA57934	Human PRO4499 cDNA
c	39	20	0.2	1977	25	ABX98404	Human cDNA encodin
c	40	20	0.2	1977	25	ABX98906	Novel human secret
c	41	20	0.2	1977	25	ACA05951	Human secreted/tra
c	42	20	0.2	1977	25	ABX97995	Human PRO polynucl
c	43	20	0.2	1977	25	ABX78779	Human PRO polynucl
c	44	20	0.2	1977	25	ABX75792	Human cDNA encodin
c	45	20	0.2	1977	25	ABX76997	Human PRO polynucl
c	46	20	0.2	1977	25	ABX16837	Human cDNA encodin
c	47	20	0.2	1981	21	AAZ51194	Human ischaemic he
	48	20	0.2	1999	22	ABA09180	Human secreted pro
	49	20	0.2	1999	22	AAK52615	Human polynucleoti
c	50	20	0.2	2003	22	AAH99744	Human protein enco
c	51	20	0.2	2008	21	AAZ51193	Human ischaemic he
c	52	20	0.2	2017	20	AAV08830	Gene No. 20 encodi
c	53	20	0.2	2031	21	AAZ56764	Human transmembran
c	54	20	0.2	2337	22	AAK51631	Human polynucleoti
c	55	20	0.2	9238	24	ABK28365	DNA transcription
c	56	20	0.2	11131	24	ABL92227	Chemically treated
c	57	20	0.2	11131	24	ABL49328	Human polynucleoti
c	58	20	0.2	11131	24	ABL32753	Human immune syste
c	59	20	0.2	11327	22	AAS46431	Tumour suppressor
	60	20	0.2	32392	24	ABL56203	AmEPV genome fragm
	61	20	0.2	35100	20	AAV73803	KSHV LUR DNA (nucl
	62	20	0.2	137507	19	AAV19941	KSHV long unique c
	63	20	0.2	1230025	20	AAX91990	Nucleotide sequenc
c	64	19	0.2	221	25	ABX40498	Bovine EST associa
	65	19	0.2	268	24	ABL84861	Human ovarian canc
	66	19	0.2	284	25	ABX86121	Corn ear-derived p
	67	19	0.2	292	25	ABX81628	Corn ear-derived p
	68	19	0.2	305	22	AAK89762	Human digestive sy
	69	19	0.2	305	22	AAK89766	Human digestive sy
	70	19	0.2	375	24	ABN20967	Human ORFX polynuc
c	71	19	0.2	387	25	ABX40398	Bovine EST associa
c	72	19	0.2	401	25	ABX18291	Human GDP-mannose
c	73	19	0.2	564	22	AAH10172	Human cDNA clone (
c	74	19	0.2	629	24	ABQ26910	Oligonucleotide fo
	75	19	0.2	629	24	ABQ26911	Oligonucleotide fo
c	76	19	0.2	719	22	AAI58779	Human polynucleoti
c	77	19	0.2	816	22	AAH06747	Human cDNA clone (
	78	19	0.2	818	21	AAC76215	Human ORFX ORF1770
	79	19	0.2	888	25	ABZ39135	N. gonorrhoeae nuc
	80	19	0.2	900	21	AAA65435	Porcine BAC-PIGF2-
c	81	19	0.2	948	25	ABX76413	Lung cancer-associ

	82	19	0.2	989	17	AAT30244	Cotton fibre cell-
	83	19	0.2	989	17	AAT13038	Cotton fibre-speci
	84	19	0.2	989	18	AAT70045	Cotton fibre speci
	85	19	0.2	989	18	AAT62614	Cotton fibre speci
	86	19	0.2	989	21	AAZ35548	cDNA sequence a co
c	87	19	0.2	1056	21	AAA59190	Endogenetic retrov
c	88	19	0.2	1186	20	AAX29726	Insert from pET21C
c	89	19	0.2	1247	20	AAX29725	Insert from pET28C
	90	19	0.2	1287	25	ABX71022	Novel human cDNA s
	91	19	0.2	1295	21	AAC50413	Arabidopsis thalia
	92	19	0.2	1298	21	AAC37673	Arabidopsis thalia
c	93	19	0.2	1400	20	AAX29772	L.lactis HsdS gene
	94	19	0.2	1408	21	AAC50427	Arabidopsis thalia
c	95	19	0.2	1412	13	AAQ32365	MAGE-9 genomic DNA
c	96	19	0.2	1412	15	AAQ72490	Tumour rejection a
c	97	19	0.2	1412	16	AAT01169	MAGE-9 gene. Homo
c	98	19	0.2	1412	20	AAX84125	MAGE-9 gene. Homo
	99	19	0.2	1412	21	AAC40335	Arabidopsis thalia
	100	19	0.2	1456	21	AAA64206	Complete sequence
c	101	19	0.2	1511	19	AAV43228	Multiple sclerosis
c	102	19	0.2	1511	20	AAX29706	Clone CL2 from MSR
c	103	19	0.2	1511	21	AAA59188	Endogenetic retrov
	104	19	0.2	1536	23	AAS72312	DNA encoding novel
c	105	19	0.2	1678	24	ABA97429	Gentiana triflora
c	106	19	0.2	1752	22	AAF75199	3md3 DNA. Homo sa
c	107	19	0.2	1831	21	AAA64634	cDNA encoding huma
c	108	19	0.2	1858	24	ABK62090	Human cDNA encodin
c	109	19	0.2	1874	21	AAZ48825	Full length Human
c	110	19	0.2	1967	25	ABZ81798	Mouse notch ligand
c	111	19	0.2	1967	25	AAD52528	Human notch ligand
c	112	19	0.2	1967	25	ABX75301	Human notch ligand
c	113	19	0.2	1983	22	AAH18571	Human cDNA sequenc
c	114	19	0.2	2003	22	AAK94621	Human full-length
c	115	19	0.2	2009	21	AAA59189	Endogenetic retrov
	116	19	0.2	2036	24	ABK99863	Babesia cDNA seque
	117	19	0.2	2043	23	ABL24009	Drosophila melanog
	118	19	0.2	2044	22	AAH13813	Human cDNA sequenc
c	119	19	0.2	2052	21	AAA63825	Nucleotide sequenc
	120	19	0.2	2115	23	AAS77887	DNA encoding novel
	121	19	0.2	2167	13	AAQ22484	gro es el operon w
	122	19	0.2	2274	23	ABL22784	Drosophila melanog
	123	19	0.2	2654	23	ABL22760	Drosophila melanog
	124	19	0.2	2668	13	AAQ22485	groEL-1 gene. Str
c	125	19	0.2	2910	23	AAS82418	DNA encoding novel
c	126	19	0.2	2910	23	AAS93250	DNA encoding novel
c	127	19	0.2	2931	19	ABQ76209	Human tumour antig
c	128	19	0.2	2931	25	ABX76208	Lung cancer-associ
	129	19	0.2	3230	22	AAS30094	Human lung antigen
	130	19	0.2	4386	21	AAZ89047	Human nibrin DNA.
	131	19	0.2	4386	21	AAZ34997	NBS1 gene associat
	132	19	0.2	4412	25	ABZ72011	Human nibrin (NBS)
	133	19	0.2	4778	23	ABL24008	Drosophila melanog
	134	19	0.2	5084	23	ABL13654	Drosophila melanog
c	135	19	0.2	5352	25	ACC50996	Human bladder canc
	136	19	0.2	5445	22	AAS46595	Tumour suppressor
c	137	19	0.2	5667	22	AAS28999	Genomic sequence #
c	138	19	0.2	5667	22	AAS31658	Genomic sequence #
c	139	19	0.2	5667	24	ABT07834	Novel human nuclei
c	140	19	0.2	5919	22	AAS46349	Tumour suppressor
c	141	19	0.2	5919	24	ABL70213	Chemically treated
c	142	19	0.2	5919	24	ABL32732	Human immune syste
c	143	19	0.2	5919	24	AAS61168	Human gene regulat
c	144	19	0.2	5919	24	ABK31258	Signal transductio
c	145	19	0.2	6191	24	ABN80160	Human chemically m
c	146	19	0.2	6191	24	ABL70281	Chemically treated
c	147	19	0.2	6191	24	ABL33216	Human immune syste
c	148	19	0.2	6191	24	ABK31306	Signal transductio
c	149	19	0.2	7118	22	AAS46381	Tumour suppressor
c	150	19	0.2	7118	24	ABL32830	Human immune syste
	151	19	0.2	7333	23	ABL25676	Drosophila melanog
c	152	19	0.2	7455	24	ABL33758	Human immune syste
	153	19	0.2	7965	24	AAI69347	C. albicans contig
c	154	19	0.2	8910	24	ABN80132	Human chemically m
c	155	19	0.2	9477	23	ABL14686	Drosophila melanog
	156	19	0.2	10000	21	AAA10667	Barley Rar1 resist
	157	19	0.2	10202	23	ABL02571	Drosophila melanog
c	158	19	0.2	13449	24	ABL33384	Human immune syste
	159	19	0.2	13508	18	AAV74431	Staphylococcus aur
	160	19	0.2	14738	22	AAL02949	Human reproductive
	161	19	0.2	14738	22	AAL02951	Human reproductive
c	162	19	0.2	15399	24	ABL33514	Human immune syste
	163	19	0.2	15405	23	ABL03146	Drosophila melanog
c	164	19	0.2	18855	24	ABL32611	Human immune syste
c	165	19	0.2	19603	23	ABL02570	Drosophila melanog
	166	19	0.2	23695	24	ABQ66982	Human angiogenesis
	167	19	0.2	33718	22	AAK65257	Human immune/haema
	168	19	0.2	33718	22	AAK86411	Human immune/haema
	169	19	0.2	65921	21	AAZ89046	Human nibrin DNA.
c	170	19	0.2	75384	22	AAK85590	Human immune/haema
	171	19	0.2	222930	24	ABK84349	Human cDNA differe
c	172	19	0.2	341511	24	ABS55200	Genomic DNA encodi
	173	19	0.2	349980	24	ABQ81848	Bifidobacterium lo
c	174	19	0.2	506286	24	ABS55320	Human transporter
	175	19	0.2	513445	22	AAI61373	Soybean 318013 reg
c	176	19	0.2	640681	24	ABA92787	Buchnera sp. genom
	177	19	0.2	684707	24	ABQ67196	Listeria innocua c
	178	19	0.2	1082138	21	AAF22305	Arabidopsis thalia
	179	19	0.2	3011208	24	ABQ69245	Listeria innocua D
c	180	18	0.2	36	21	AAA15143	Probe for DNA enco
	181	18	0.2	111	22	ABA76541	Human foetal liver
	182	18	0.2	111	22	ABA41049	Probe #19515 for g
	183	18	0.2	111	22	AAK25179	Human brain expres
	184	18	0.2	111	22	AAK51184	Human bone marrow
	185	18	0.2	111	22	AAI28194	Probe #18127 for g
	186	18	0.2	111	22	AAI57237	Probe #25923 used
	187	18	0.2	111	23	ABS50750	Human liver single
	188	18	0.2	111	24	ABS24715	Human genome-deriv
c	189	18	0.2	121	20	AAH86310	Human single nucle
c	190	18	0.2	127	24	ABL72211	Corn tassel-derive
c	191	18	0.2	172	24	ABS23894	Human genome-deriv
	192	18	0.2	222	22	AAL00427	Human reproductive
	193	18	0.2	269	24	ABL71665	Corn tassel-derive
c	194	18	0.2	279	24	ABL71121	Corn tassel-derive
	195	18	0.2	291	22	ABA73860	Human foetal liver
	196	18	0.2	291	22	ABA39006	Probe #17472 for g
	197	18	0.2	291	22	AAK22313	Human brain expres
	198	18	0.2	291	22	AAK48478	Human bone marrow
	199	18	0.2	291	22	AAI26246	Probe #16179 for g
	200	18	0.2	291	22	AAI54306	Probe #22992 used
	201	18	0.2	291	23	ABS48172	Human liver single
	202	18	0.2	291	24	ABS22278	Human genome-deriv
	203	18	0.2	295	22	AAL18260	Human breast cance
	204	18	0.2	300	21	AAA01511	Human colon cancer
c	205	18	0.2	300	23	ABV61275	Human prostate exp
	206	18	0.2	311	20	AAZ40537	Human STLK7 DNA.
	207	18	0.2	338	22	AAL08816	Human breast cance
	208	18	0.2	357	23	ABV51636	Human prostate exp
	209	18	0.2	372	21	AAC99890	Human secreted pro
	210	18	0.2	376	25	ABZ55214	Aspergillus oryzae
	211	18	0.2	376	25	ABX42317	Bovine EST associa
c	212	18	0.2	400	21	AAC09327	Human secreted pro
c	213	18	0.2	402	24	ABN96275	Gene #2773 used to
c	214	18	0.2	414	23	ABV37381	Human prostate exp
	215	18	0.2	419	23	AAS72311	DNA encoding novel
c	216	18	0.2	422	22	AAI87211	Human polynucleoti
	217	18	0.2	429	22	AAK59648	Human immune/haema
	218	18	0.2	433	25	ABX36578	Bovine EST associa
	219	18	0.2	434	24	ABL36429	Human colon tumour
	220	18	0.2	434	25	ABX35185	Bovine EST associa
	221	18	0.2	435	23	ABV15958	Human prostate exp
	222	18	0.2	437	22	AAK75804	Human immune/haema
c	223	18	0.2	440	22	AAL36018	Human musculoskele
c	224	18	0.2	440	22	AAK72991	Human immune/haema
c	225	18	0.2	440	25	ABX59006	cDNA encoding nove
	226	18	0.2	446	22	AAK92916	Human cDNA 3'-end
c	227	18	0.2	466	19	AAV65212	DNA encoding a S.

	228	18	0.2	469	22	AAK88191 Human digestive sy
c	229	18	0.2	474	25	ABX07356 S. pneumoniae type
c	230	18	0.2	488	23	ABL16517 Drosophila melanog
	231	18	0.2	493	23	ABV45757 Human prostate exp
c	232	18	0.2	503	24	ABV88113 Human colon cancer
	233	18	0.2	511	23	ABK41891 cDNA encoding nove
	234	18	0.2	519	22	ABA61361 Human foetal liver
	235	18	0.2	519	22	ABA29155 Probe #7621 for ge
	236	18	0.2	519	22	AAK09661 Human brain expres
	237	18	0.2	519	22	AAK35552 Human bone marrow
	238	18	0.2	519	22	AAI17068 Probe #7001 for ge
	239	18	0.2	519	22	AAI41267 Probe #9953 used t
	240	18	0.2	519	23	ABS35282 Human liver single
	241	18	0.2	519	24	ABS09875 Human genome-deriv
c	242	18	0.2	524	24	ABN83559 Human acyl-CoA deh
c	243	18	0.2	550	22	AAH10106 Human cDNA clone (
c	244	18	0.2	558	24	ABS11536 Human genome-deriv
c	245	18	0.2	560	22	AAH53839 S. epidermidis ope
c	246	18	0.2	564	24	ABK63159 Rat sequence diffe
	247	18	0.2	577	22	ABA64103 Human foetal liver
	248	18	0.2	577	22	ABA31262 Probe #9728 for ge
	249	18	0.2	577	22	AAK12584 Human brain expres
	250	18	0.2	577	22	AAK38300 Human bone marrow
	251	18	0.2	577	22	AAI19098 Probe #9031 for ge
	252	18	0.2	577	22	AAI44248 Probe #12934 used
	253	18	0.2	577	23	ABS37914 Human liver single
	254	18	0.2	577	24	ABS12351 Human genome-deriv
c	255	18	0.2	585	22	ABA60444 Human foetal liver
c	256	18	0.2	585	22	AAK08723 Human brain expres
c	257	18	0.2	585	22	AAK34610 Human bone marrow
c	258	18	0.2	585	22	AAI40329 Probe #9015 used t
c	259	18	0.2	585	23	ABS34384 Human liver single
c	260	18	0.2	585	24	ABS09234 Human genome-deriv
c	261	18	0.2	595	23	ABV51243 Human prostate exp
c	262	18	0.2	615	22	AAF71466 Corynebacterium gl
c	263	18	0.2	615	22	AAC85429 hFGF-8 coding sequ
c	264	18	0.2	619	23	ABV54871 Human prostate exp
c	265	18	0.2	645	22	AAH44396 Human FGF-8 nucleo
c	266	18	0.2	645	22	AAH48265 Heart muscle cell
c	267	18	0.2	645	24	ABN92523 Staphylococcus epi
c	268	18	0.2	648	21	AAA09456 Human fibroblast g
c	269	18	0.2	648	22	AAC85430 hFGF-8 coding sequ
c	270	18	0.2	648	24	AAD43987 Human FGF8b (fibro
c	271	18	0.2	649	22	AAH84125 Papio hamadryas ol
c	272	18	0.2	649	24	ABZ35776 Human FGF8 polynuc
c	273	18	0.2	649	24	ABX10019 Human FGF8 DNA fra
c	274	18	0.2	649	24	ABV78200 Human FGF8 DNA SEQ
c	275	18	0.2	649	24	ABL91741 Human polynucleoti
c	276	18	0.2	650	22	AAH84127 Papio hamadryas ol
c	277	18	0.2	663	22	AAS30984 Human cDNA encodin
c	278	18	0.2	672	22	AAH53842 S. epidermidis ope
	279	18	0.2	696	24	ABN91524 Staphylococcus epi
c	280	18	0.2	702	22	AAC85431 hFGF-8 coding sequ
	281	18	0.2	715	20	AAX08422 G protein gene fra
c	282	18	0.2	726	24	ABK78747 Bacillus clausii g
c	283	18	0.2	732	22	AAC85432 hFGF-8 coding sequ
	284	18	0.2	744	23	ABV06879 Human prostate exp
c	285	18	0.2	819	24	ABQ14224 Oligonucleotide fo
	286	18	0.2	819	24	ABQ14225 Oligonucleotide fo
	287	18	0.2	848	24	ABQ73737 Human colon specif
	288	18	0.2	920	14	AAQ45686 Respiratory syncit
	289	18	0.2	920	20	AAX08421 Membrane bound G p
c	290	18	0.2	997	14	AAQ48024 AIGF gene #1. Hom
c	291	18	0.2	997	22	AAD16024 Mouse fibroblast g
c	292	18	0.2	1059	22	AAH65335 C glutamicum codin
	293	18	0.2	1062	24	ABZ13834 Arabidopsis thalia
c	294	18	0.2	1062	24	ABL32152 Human immune syste
c	295	18	0.2	1100	20	AAX60110 DNA encoding a for
	296	18	0.2	1111	24	ABQ73738 Human colon specif
c	297	18	0.2	1164	24	ABQ70288 Listeria monocytog
c	298	18	0.2	1178	14	AAQ48025 AIGF gene #2. Hom
c	299	18	0.2	1187	24	ABN83323 Murine renin recep
	300	18	0.2	1194	23	AAS67475 DNA encoding novel

ALIGNMENTS

```
RESULT 1
AAI64291
ID   AAI64291 standard; DNA; 128139 BP.
XX
AC   AAI64291;
XX
DT   22-APR-2002  (first entry)
XX
DE   RRV genome nucleotide sequence.
XX
KW   RRV; rhesus rhadinovirus; japanese macaque virus; multiple sclerosis;
KW   JMHV; cytostatic; antiasthmatic; antiallergic; dermatological;
KW   vulnerary; gene therapy; leucopenia; thrombocytopaenia;
KW   inflammatory disease; asthma; allergy; dermatitis; virus; ds.
XX
OS   Macaca mulatta rhadinovirus 17577.
XX
FH   Key             Location/Qualifiers
FT   CDS             1353..2674
FT                   /*tag= a
FT                   /product= "RRV R1"
FT   CDS             complement (2692..3258)
FT                   /*tag= b
FT                   /product= "dihydrofolate reductase"
FT                   /label= RRV_ORF2
FT                   /note= "has similarity to Kaposi's sarcoma-associated
FT                         virus (KSHV) open reading frame (ORF) 2"
FT   CDS             3676..5613
FT                   /*tag= c
FT                   /product= "complement binding protein"
FT                   /label= RRV_ORF4
FT                   /note= "has similarity to KSHV ORF4"
FT   CDS             6045..9443
FT                   /*tag= d
FT                   /product= "ssDNA binding protein"
FT                   /label= RRV_ORF6
FT                   /note= "has similarity to KSHV ORF6"
FT   CDS             9468..11528
FT                   /*tag= e
FT                   /product= "transport protein"
FT                   /label= RRV_ORF7
FT                   /note= "has similarity to KSHV ORF7"
FT   CDS             11515..14004
FT                   /*tag= f
FT                   /product= "glycoprotein B"
FT                   /label= RRV_ORF8
FT                   /note= "has similarity to KSHV ORF8"
FT   CDS             14122..17166
FT                   /*tag= g
FT                   /product= "DNA polymerase protein"
FT                   /label= RRV_ORF9
FT                   /note= "has similarity to KSHV ORF9"
FT   CDS             17261..18511
FT                   /*tag= h
FT                   /label= RRV_ORF10
FT                   /note= "has similarity to KSHV ORF10"
FT   CDS             18520..19749
FT                   /*tag= i
FT                   /label= RRV_ORF11
FT                   /note= "has similarity to KSHV ORF11"
FT   CDS             complement (19921..20544)
FT                   /*tag= j
FT                   /product= "RRV R2"
FT                   /label= RRV_ORF
FT                   /note= "has similarity to KSHV interleukin (IL)-6 gene"
FT   CDS             complement (20777..21778)
FT                   /*tag= k
FT                   /product= "thymidylate synthase"
```

```
FT                   /label= RRV_ORF70
FT                   /note= "has similarity to KSHV ORF70"
FT   CDS             complement (22245..22592)
FT                   /*tag=  l
FT                   /product= "RRV R3"
FT                   /note= "has similarity to KSHV K4 viral MIP gene"
FT   CDS             26846..27409
FT                   /*tag=  m
FT                   /product= "Bcl2-homologue"
FT                   /label= RRV_ORF16
FT                   /note= "has similarity to KSHV ORF16"
FT   CDS             complememt (27515..29125)
FT                   /*tag=  n
FT                   /label= RRV_ORF17
FT                   /note= "has similarity to KSHV ORF17"
FT   CDS             28998..29897
FT                   /*tag=  o
FT                   /label= RRV_ORF18
FT                   /note= "has similarity to KSHV ORF18"
FT   CDS             complement (29905..31548)
FT                   /*tag=  p
FT                   /product= "tegument protein
FT                   /label= RRV_ORF19
FT                   /note= "has similarity to KSHV ORF19"
FT   CDS             complement (31043..32095)
FT                   /*tag=  q
FT                   /label= RRV_ORF20
FT                   /note= "has similarity to KSHV ORF20"
FT   CDS             32094..33767
FT                   /*tag=  r
FT                   /product= "thymidine kinase"
FT                   /label= RRV_ORF21
FT                   /note= "has similarity to KSHV ORF21"
FT   CDS             33754..35868
FT                   /*tag=  s
FT                   /product= "glycoprotein H"
FT                   /label= RRV_ORF22
FT                   /note= "has similarity to KSHV ORF22"
FT   CDS             complement (35865..37073)
FT                   /*tag=  t
FT                   /label= RRV_ORF23
FT                   /note= "has similarity to KSHV ORF23"
FT   CDS             complement (37123..39321)
FT                   /*tag=  u
FT                   /label= RRV_ORF24
FT                   /note= "has similarity to KSHV ORF24"
FT   CDS             39323..43459
FT                   /*tag=  v
FT                   /product= "major capsid protein"
FT                   /label= RRV_ORF25
FT                   /note= "has similarity to KSHV ORF25"
FT   CDS             43491..44408
FT                   /*tag=  w
FT                   /product= "capsid protein"
FT                   /label= RRV_ORF26
FT                   /note= "has similarity to KSHV ORF26"
FT   CDS             44433..45242
FT                   /*tag=  x
FT                   /label= RRV_ORF27
FT                   /note= "has similarity to KSHV ORF27"
FT   CDS             45408..45683
FT                   /*tag=  y
FT                   /label= RRV_ORF28
FT                   /note= "has similarity to KSHV ORF28"
FT   CDS             complement (45733..46779)
FT                   /*tag=  z
FT                   /label= RRV_ORF29b
FT                   /note= "has similarity to KSHV ORF29b"
FT   CDS             46905..47135
FT                   /*tag=  aa
FT                   /label= RRV_ORF30
FT                   /note= "has similarity to KSHV ORF30"
FT   CDS             47093..47746
FT                   /*tag=  ab
FT                   /label= RRV_ORF31
FT                   /note= "has similarity to KSHV ORF31"
FT   CDS             47683..49077
FT                   /*tag=  ac
FT                   /label= RRV_ORF32
FT                   /note= "has similarity to KSHV ORF32"
FT   CDS             49049..50059
FT                   /*tag=  ad
FT                   /label= RRV_ORF33
FT                   /note= "has similarity to KSHV ORF33"
FT   CDS             complement (49977..50960)
FT                   /*tag=  ae
FT                   /label= RRV_ORF29a
FT                   /note= "has similarity to KSHV ORF29a"
FT   CDS             50959..51942
FT                   /*tag=  af
FT                   /label= RRV_ORF34
FT                   /note= "has similarity to KSHV ORF34"
FT   CDS             51923..52372
FT                   /*tag=  ag
FT                   /label= RRV_ORF35
FT                   /note= "has similarity to KSHV ORF35"
FT   CDS             52278..53585
FT                   /*tag=  ah
FT                   /product= "kinase"
FT                   /label= RRV_ORF36
FT                   /note= "has similarity to KSHV ORF36"
FT   CDS             53566..55008
FT                   /*tag=  ai
FT                   /product= "alkaline exonuclease"
FT                   /label= RRV_ORF37
FT                   /note= "has similarity to KSHV ORF37"
FT   CDS             54963..55172
FT                   /*tag=  aj
FT                   /label= RRV_ORF38
FT                   /note= "has similarity to KSHV ORF38"
FT   CDS             complement (55255..56391)
FT                   /*tag=  ak
FT                   /product= "glycoprotein M"
FT                   /label= RRV_ORF39
FT                   /note= "has similarity to KSHV ORF39"
FT   CDS             56526..57932
FT                   /*tag=  al
FT                   /product= "helicase/primase"
FT                   /label= RRV_ORF40
FT                   /note= "has similarity to KSHV ORF40"
FT   CDS             57917..58528
FT                   /*tag=  am
FT                   /product= "helicase/primase"
FT                   /label= RRV_ORF41
FT                   /note= "has similarity to KSHV ORF41"
FT   CDS             complement (58525..59343)
FT                   /*tag=  an
FT                   /label= RRV_ORF42
FT                   /note= "has similarity to KSHV ORF42"
FT   CDS             complement (59297..61027)
FT                   /*tag=  ao
FT                   /product= "capsid protein"
FT                   /label= RRV_ORF43
FT                   /note= "has similarity to KSHV ORF43"
```

```
  Query Match             100.0%;  Score 9073;  DB 24;  Length 128139;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 9073;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGAGAGTTCCGTCGGATGGACCAAACACGTCGAACCAAATCCGGGGTTCATCTTGAAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55386 ATGGAGAGTTCCGTCGGATGGACCAAACACGTCGAACCAAATCCGGGGTTCATCTTGAAC 55445

Qy         61 ATGACGTCCGATGCCAAAGTCAGGGGTGTCGTGGATCACGTCAGTCGCCTGTCAAATATA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55446 ATGACGTCCGATGCCAAAGTCAGGGGTGTCGTGGATCACGTCAGTCGCCTGTCAAATATA 55505
```

```
Qy       121 ACTACCAGCCCACCGGAAATGGGTTGGTACGACCTGGCCTTCGATCCGGCTGAAGACTCC 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     55506 ACTACCAGCCCACCGGAAATGGGTTGGTACGACCTGGCCTTCGATCCGGCTGAAGACTCC 55565

Qy       181 GGGCCGTTCTTGCCGTTTACCGTTTATCTAATTACGGGAACTGCTGGTGCTGGGAAAAGT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     55566 GGGCCGTTCTTGCCGTTTACCGTTTATCTAATTACGGGAACTGCTGGTGCTGGGAAAAGT 55625

Qy       241 ACCAGCATATCGGCCCTGTACCAAAATTTAAACTGCCTGATCACGGGCGCGACCACCATA 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     55626 ACCAGCATATCGGCCCTGTACCAAAATTTAAACTGCCTGATCACGGGCGCGACCACCATA 55685

Qy       301 GCCGCACAGAACCTATCGCGTCGCCTAAAGACGTTCTGTCCCACGATCTTCAGCGCTTTT 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     55686 GCCGCACAGAACCTATCGCGTCGCCTAAAGACGTTCTGTCCCACGATCTTCAGCGCTTTT 55745

Qy       361 GGCTTTAAGAGCCGACACATCAATATAGCCGTCAGAAAAGCTCATCAGACCGGAGCCGTA 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     55746 GGCTTTAAGAGCCGACACATCAATATAGCCGTCAGAAAAGCTCATCAGACCGGAGCCGTA 55805

Qy       421 TCCATAGAGCAAATTCAGCAACAGGAGCTATCGAAGTATTGGCCGGTTATAGTGGACATT 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     55806 TCCATAGAGCAAATTCAGCAACAGGAGCTATCGAAGTATTGGCCGGTTATAGTGGACATT 55865

Qy       481 ATGAAAGAGGTTATGGCGAAAAAACCCAATGGCATGTACGGGACTATATCCAACGCGAAT 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     55866 ATGAAAGAGGTTATGGCGAAAAAACCCAATGGCATGTACGGGACTATATCCAACGCGAAT 55925

Qy       541 TTTGAAACCCTCTCGAGAATGACCGGACCGTGTTTATGGACTTCCAATATTATTGTAATC 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     55926 TTTGAAACCCTCTCGAGAATGACCGGACCGTGTTTATGGACTTCCAATATTATTGTAATC 55985

Qy       601 GACGAGGCCGGAACCCTGTCCTCTTACATACTCACCACCGTCGTGTTCTTTTACTGGTTC 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     55986 GACGAGGCCGGAACCCTGTCCTCTTACATACTCACCACCGTCGTGTTCTTTTACTGGTTC 56045

Qy       661 CTAAACAGCTGGCTAAATACCCCTCTTTACCGCCAGGGGGCGGTTCCGTGCATAGTATGC 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     56046 CTAAACAGCTGGCTAAATACCCCTCTTTACCGCCAGGGGGCGGTTCCGTGCATAGTATGC 56105

Qy       721 GTCGGTTCGCCAACGCAGACAAACGCGTTCCAGTCAACTTACAACCACGGGACGCAAAAG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     56106 GTCGGTTCGCCAACGCAGACAAACGCGTTCCAGTCAACTTACAACCACGGGACGCAAAAG 56165

Qy       781 ACGGAGATATCGTCGTGCGAGAACATCCTAACATTCATGATAGGGAAGAAGGTCGTATCT 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     56166 ACGGAGATATCGTCGTGCGAGAACATCCTAACATTCATGATAGGGAAGAAGGTCGTATCT 56225

Qy       841 GAGTACGTACACTTGGAGAGAAACTGGGCGCTGTTTATAAACAACAAGCGCTGCACCGAT 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     56226 GAGTACGTACACTTGGAGAGAAACTGGGCGCTGTTTATAAACAACAAGCGCTGCACCGAT 56285

Qy       901 CTGCAGTTCGGACACCTTCTAAAAATTTTAGAGTATAATCTTCCCATTCCTGACGAAGTC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     56286 CTGCAGTTCGGACACCTTCTAAAAATTTTAGAGTATAATCTTCCCATTCCTGACGAAGTC 56345

Qy       961 ATGAGTTACGTAGACAGATTTGTCGTTCCTAAAAGTAAGATTATGGATCCTTTAGAATAC 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     56346 ATGAGTTACGTAGACAGATTTGTCGTTCCTAAAAGTAAGATTATGGATCCTTTAGAATAC 56405

Qy      1021 ATTGGCTGGACCCGACTCTTTTTGTCACATAGCGAGGTAAAGGCGTATCTAACAAACCTA 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     56406 ATTGGCTGGACCCGACTCTTTTTGTCACATAGCGAGGTAAAGGCGTATCTAACAAACCTA 56465

Qy      1081 CACACATGTCTAACGCTAGGGGGCGATACCAGGGACACGAAGCTCTTTACCTGTCCCGTG 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     56466 CACACATGTCTAACGCTAGGGGGCGATACCAGGGACACGAAGCTCTTTACCTGTCCCGTG 56525

Qy      1141 GTGTGCGAGGTGTTTGTGAAGCCGTTTGAGGAATACAAGCGGGCCGTCAACCTCACCAAC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     56526 GTGTGCGAGGTGTTTGTGAAGCCGTTTGAGGAATACAAGCGGGCCGTCAACCTCACCAAC 56585

Qy      1201 CTCACCGTGACCGAGTGGGTGACAAAAAACCTCTTTAAGTTAAGTAATTATTCGCAGTTT 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     56586 CTCACCGTGACCGAGTGGGTGACAAAAAACCTCTTTAAGTTAAGTAATTATTCGCAGTTT 56645

Qy      1261 GTGGACCAGGACATGTCCATAGTCGCCACGGAATCCACTGAACGTTCGACCCAGGTTACC 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     56646 GTGGACCAGGACATGTCCATAGTCGCCACGGAATCCACTGAACGTTCGACCCAGGTTACC 56705

Qy      1321 TTTATCACCAAGTTTGTTAAAAACAGCCACGTGTCTTTAAATGGAAAAACAAAAAAATGC 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     56706 TTTATCACCAAGTTTGTTAAAAACAGCCACGTGTCTTTAAATGGAAAAACAAAAAAATGC 56765

Qy      1381 ATATGTGGGTTTCAGGGTACGTACTTTGAGTTCAAAAGAATCCTAGACAGCGAACTCTTC 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     56766 ATATGTGGGTTTCAGGGTACGTACTTTGAGTTCAAAAGAATCCTAGACAGCGAACTCTTC 56825

Qy      1441 GTGGAGACGCATTCGCAAGACCGTCCGGAATACGTATACGGTTTCTTAAACACGCTACTG 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     56826 GTGGAGACGCATTCGCAAGACCGTCCGGAATACGTATACGGTTTCTTAAACACGCTACTG 56885

Qy      1501 TACAACGCCATGTACTCGTTTCACGCGTACGGCGTGACTAGGTCACACGAGAAATATCTG 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     56886 TACAACGCCATGTACTCGTTTCACGCGTACGGCGTGACTAGGTCACACGAGAAATATCTG 56945

Qy      1561 CAAGACCTAAAATTTGCACCCCTCCCGGCCGCTCTGGCAACCGGGCGTGTAGACCTTCAA 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     56946 CAAGACCTAAAATTTGCACCCCTCCCGGCCGCTCTGGCAACCGGGCGTGTAGACCTTCAA 57005

Qy      1621 ACGGTTCGTGAAGAGTTAAACCTGGAAGACGACATCTTCTACCACGTGTGCAGTCCCCCG 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     57006 ACGGTTCGTGAAGAGTTAAACCTGGAAGACGACATCTTCTACCACGTGTGCAGTCCCCCG 57065

Qy      1681 CCACCCGCGGGTATCACCTCCCTCCAGGTTTTGGTCGACACGTACTGCGCCCTAAAGGAC 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     57066 CCACCCGCGGGTATCACCTCCCTCCAGGTTTTGGTCGACACGTACTGCGCCCTAAAGGAC 57125

Qy      1741 GTGTTCGCCTCCAGAATAAAGGTGGCGTGTCGCTGGTTTGGCGGGGAGTTTGAGAAGGAA 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     57126 GTGTTCGCCTCCAGAATAAAGGTGGCGTGTCGCTGGTTTGGCGGGGAGTTTGAGAAGGAA 57185

Qy      1801 ACGTTTTCCGCGTTTACGGTTAACATGGTCGTAAGGGACGGAGTTGACTTTGTCTCCCCT 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     57186 ACGTTTTCCGCGTTTACGGTTAACATGGTCGTAAGGGACGGAGTTGACTTTGTCTCCCCT 57245

Qy      1861 TCAGAACGTCTCAACGGGCTGTTGGCGTTTGCATCGACCGTTGAATCGTATAAAATTAAG 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     57246 TCAGAACGTCTCAACGGGCTGTTGGCGTTTGCATCGACCGTTGAATCGTATAAAATTAAG 57305

Qy      1921 GGGTACACGTTTTTACCGGTAGCGTTCGGTCGCTGTCAGGGTTTGCCACTCAGTGATGAC 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     57306 GGGTACACGTTTTTACCGGTAGCGTTCGGTCGCTGTCAGGGTTTGCCACTCAGTGATGAC 57365

Qy      1981 CTCAGGAAGAAGATGCCCTCCCTGGTCGTGCAGGACTCTAGCGGTTTTATCGCGTGCCTA 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     57366 CTCAGGAAGAAGATGCCCTCCCTGGTCGTGCAGGACTCTAGCGGTTTTATCGCGTGCCTA 57425

Qy      2041 GAGAATAACATAACCAAATTGACCGAAACCATGGAGGACGGGAGCGTTTTCCAAGTGTGC 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     57426 GAGAATAACATAACCAAATTGACCGAAACCATGGAGGACGGGAGCGTTTTCCAAGTGTGC 57485

Qy      2101 TGTGCGGGGGACTATGGGGTCAGCTCAAATTTAGCCATGACCATCGTAAAGGCACAGGGA 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     57486 TGTGCGGGGGACTATGGGGTCAGCTCAAATTTAGCCATGACCATCGTAAAGGCACAGGGA 57545

Qy      2161 ATGTCGTTGGAGCGAGTAGCCGTAGTATTTGGGTCCCACAAGAACGTCCAGACAAGCCAC 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     57546 ATGTCGTTGGAGCGAGTAGCCGTAGTATTTGGGTCCCACAAGAACGTCCAGACAAGCCAC 57605

Qy      2221 GTGTATGTAGCAATATCAAGGGCTGTTAACTCAAATTATTTGGTCATGGACAGCAACCCC 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     57606 GTGTATGTAGCAATATCAAGGGCTGTTAACTCAAATTATTTGGTCATGGACAGCAACCCC 57665

Qy      2281 CTTAAAACCCTCCTCAGAGAACCAGTCGATAACACCTCCGCCAAGCATATAGTCCGCGCC 2340
```

```
Db      57666 CTTAAAACCCTCCTCAGAGAACCAGTCGATAACACCTCCGCCAAGCATATAGTCCGCGCC 57725
Qy       2341 CTCCACAACCCAAACACAACCCTCATCTACTAAATAAATAAAAACAAGCAAATGGAAAAC  2400
Db      57726 CTCCACAACCCAAACACAACCCTCATCTACTAAATAAATAAAAACAAGCAAATGGAAAAC 57785
Qy       2401 ATTGTGTTTTTATTCAGTCCAACCACGGCCACGGATAGTTGTCATTTCCACACACCGGGG  2460
Db      57786 ATTGTGTTTTTATTCAGTCCAACCACGGCCACGGATAGTTGTCATTTCCACACACCGGGG 57845
Qy       2461 GGGTGGCCGCCATAGTTTGACGACCAGAAACGCCCGGGCCCTCTGGGTTAGAAGTAGCCC  2520
Db      57846 GGGTGGCCGCCATAGTTTGACGACCAGAAACGCCCGGGCCCTCTGGGTTAGAAGTAGCCC 57905
Qy       2521 CGCCCCCCGTGGTCGAGGTCATACGTCTCTTAGGGCCCTGGGTTGTACCAACATAAATCA  2580
Db      57906 CGCCCCCCGTGGTCGAGGTCATACGTCTCTTAGGGCCCTGGGTTGTACCAACATAAATCA 57965
Qy       2581 CGTCGTCATCACCGTTGCAGTCCCTGCGTGACGCGCGCCTCTTGCAGGCGGCACGAACAC  2640
Db      57966 CGTCGTCATCACCGTTGCAGTCCCTGCGTGACGCGCGCCTCTTGCAGGCGGCACGAACAC 58025
Qy       2641 CGCCGCTTGTAACGTGATCTGCATCGTCGTCGCTGTCTGAAGATGAACACAGGTCAATCA  2700
Db      58026 CGCCGCTTGTAACGTGATCTGCATCGTCGTCGCTGTCTGAAGATGAACACAGGTCAATCA 58085
Qy       2701 CATCAGTGGCGCCCGTGCCCTCATCGGCGTCTGATTCGTCCCATGTCTCAACAGAGCCTC  2760
Db      58086 CATCAGTGGCGCCCGTGCCCTCATCGGCGTCTGATTCGTCCCATGTCTCAACAGAGCCTC 58145
Qy       2761 GGTCGTCCACTTCGTCGTCATCTGAAATAATAATGGTCTCTTGTCCTGCAGTTGATCCAG  2820
Db      58146 GGTCGTCCACTTCGTCGTCATCTGAAATAATAATGGTCTCTTGTCCTGCAGTTGATCCAG 58205
Qy       2821 CAACAACAGGTTCTTCTTCATCGGAAGACCATCCGCCCTCTGCCGGTGATCTAGATATAT  2880
Db      58206 CAACAACAGGTTCTTCTTCATCGGAAGACCATCCGCCCTCTGCCGGTGATCTAGATATAT 58265
Qy       2881 CACCTGACCAGGCGGGCGAGGCCCGTGGCGAAAAGCCGTCAGCTTCTCCGCCACTTTCAT  2940
Db      58266 CACCTGACCAGGCGGGCGAGGCCCGTGGCGAAAAGCCGTCAGCTTCTCCGCCACTTTCAT 58325
Qy       2941 CAGTGTCCTCGTAGTCATCTATAAAAAAATCGTCTTCGCTGTCTGATTCAGATGACTGGG  3000
Db      58326 CAGTGTCCTCGTAGTCATCTATAAAAAAATCGTCTTCGCTGTCTGATTCAGATGACTGGG 58385
Qy       3001 GTGCGCGGTGTTGGCCAGTAACCACCGCAACGGGCCGCGGACCACCAAACGCGGTCAGTG  3060
Db      58386 GTGCGCGGTGTTGGCCAGTAACCACCGCAACGGGCCGCGGACCACCAAACGCGGTCAGTG 58445
Qy       3061 GCGCGGGCGCGCTCGGTGGTGTGGTGGCACCTGTCGAAGTAATGACGGTAGGTAGCCGGC  3120
Db      58446 GCGCGGGCGCGCTCGGTGGTGTGGTGGCACCTGTCGAAGTAATGACGGTAGGTAGCCGGC 58505
Qy       3121 CGCTCTGGTTAAAGGACTCACCGTCCCGTGCGTGCGGATCTATTCTAGGTGGGGAGGAGA  3180
Db      58506 CGCTCTGGTTAAAGGACTCACCGTCCCGTGCGTGCGGATCTATTCTAGGTGGGGAGGAGA 58565
Qy       3181 CCGGCGGTTCGTCTTCTTCATCCACATCACTGTAGTCGCTGTTCTCCGAATGAGAGTCGG  3240
Db      58566 CCGGCGGTTCGTCTTCTTCATCCACATCACTGTAGTCGCTGTTCTCCGAATGAGAGTCGG 58625
Qy       3241 TTTCGTCTGGCGTTTTTGGAGTAGTGGGAGGGGAATCTTGGTCAAAAACATCCTCAGGTG  3300
Db      58626 TTTCGTCTGGCGTTTTTGGAGTAGTGGGAGGGGAATCTTGGTCAAAAACATCCTCAGGTG 58685
Qy       3301 CAGTCATGTCAATCACTGGGGGGCCTCCGTACCCGGGCGGTACCGGCAACCCCGTACGGG  3360
Db      58686 CAGTCATGTCAATCACTGGGGGGCCTCCGTACCCGGGCGGTACCGGCAACCCCGTACGGG 58745
Qy       3361 ACCCGGCGAACAAAAAACGTCTTGTTCTCTTTTTCCTAGGTGCCCCGGGAATCGGCAGCA  3420
Db      58746 ACCCGGCGAACAAAAAACGTCTTGTTCTCTTTTTCCTAGGTGCCCCGGGAATCGGCAGCA 58805
Qy       3421 TCCTGGGAGTGGCCGGCGGGGTCCTTGACGGGTCGGATAAGAACATAGCCATGGCCGAAA  3480
Db      58806 TCCTGGGAGTGGCCGGCGGGGTCCTTGACGGGTCGGATAAGAACATAGCCATGGCCGAAA 58865
Qy       3481 CGTTCACCTGTAAAACGCACTGCCTATCCCCGATCAGTTAATATTCCAGTCAACGGCACC  3540
Db      58866 CGTTCACCTGTAAAACGCACTGCCTATCCCCGATCAGTTAATATTCCAGTCAACGGCACC 58925
Qy       3541 CCTCCGATTCTGAACTAGATAGTCATTCGCAAGTTTAAAATGGTTGCAGCCTAGAAACTG  3600
Db      58926 CCTCCGATTCTGAACTAGATAGTCATTCGCAAGTTTAAAATGGTTGCAGCCTAGAAACTG 58985
Qy       3601 CGGCCAGGGAGAACCGGTTGCCGCGTGGGCGGTAGCCAGGGGAGACGGATGGCGCGCCTT  3660
Db      58986 CGGCCAGGGAGAACCGGTTGCCGCGTGGGCGGTAGCCAGGGGAGACGGATGGCGCGCCTT 59045
Qy       3661 GAGGACGAGATGTCTCTGTCCGTTTATCAGCGTCGCCTTGTCTATAGCCTTTCTTCCCCA  3720
Db      59046 GAGGACGAGATGTCTCTGTCCGTTTATCAGCGTCGCCTTGTCTATAGCCTTTCTTCCCCA 59105
Qy       3721 CAGCATAAAAACGCACCTCTGAAGCTTGGCAGAAAGGCAGCTTATTATGTAGCTGGTGAA  3780
Db      59106 CAGCATAAAAACGCACCTCTGAAGCTTGGCAGAAAGGCAGCTTATTATGTAGCTGGTGAA 59165
Qy       3781 CCAATCCCAGCCAAGGTTGGCGTGTGACCCCGCCTTCCCTCTTTCCACCGTCAGAATGGT  3840
Db      59166 CCAATCCCAGCCAAGGTTGGCGTGTGACCCCGCCTTCCCTCTTTCCACCGTCAGAATGGT 59225
Qy       3841 GTTTAAAAGCAGAACTCCCCGTTTTGCCCAACAGTCTAAGCACCCGTGAGAAGGAGCGGT  3900
Db      59226 GTTTAAAAGCAGAACTCCCCGTTTTGCCCAACAGTCTAAGCACCCGTGAGAAGGAGCGGT 59285
Qy       3901 GAACCCAGGTACAGTATTGGCTATCTCTTTAAAAATATTTTTGAGGCTTGGAGGTATACT  3960
Db      59286 GAACCCAGGTACAGTATTGGCTATCTCTTTAAAAATATTTTTGAGGCTTGGAGGTATACT 59345
Qy       3961 GTAGTCCGGAGCAACGCTAAAGGCTAGTCCGGTGGCTTGACCGCGATGGTAAGGATCCTG  4020
Db      59346 GTAGTCCGGAGCAACGCTAAAGGCTAGTCCGGTGGCTTGACCGCGATGGTAAGGATCCTG 59405
Qy       4021 GCCAAGGATCACGACTTTAATATCCTCCGGTTCGCAACAATACGACCACCACATAATTCT  4080
Db      59406 GCCAAGGATCACGACTTTAATATCCTCCGGTTCGCAACAATACGACCACCACATAATTCT 59465
Qy       4081 ATCTATCGGTGGATAAATCACGGTTACGTTGCTCATATCCATAACGCGCTTCAGCAGAGC  4140
Db      59466 ATCTATCGGTGGATAAATCACGGTTACGTTGCTCATATCCATAACGCGCTTCAGCAGAGC 59525
Qy       4141 CGCGAGCTTCTGTTTTAAAAATGGAGAAAGATTTAAAAATTCAAGCCATGAGTCGCTAAG  4200
Db      59526 CGCGAGCTTCTGTTTTAAAAATGGAGAAAGATTTAAAAATTCAAGCCATGAGTCGCTAAG 59585
Qy       4201 TAGCAACGTTTGTGTGGAGGGTTCTTCTAACACTTCTGGTGACATTTTACTCCAAACTAT  4260
Db      59586 TAGCAACGTTTGTGTGGAGGGTTCTTCTAACACTTCTGGTGACATTTTACTCCAAACTAT 59645
Qy       4261 TGTTTTAAGCCAACCCTCCATAAATCCCCGCGTCTGGTAATTACTTTACGTATACGTGAA  4320
Db      59646 TGTTTTAAGCCAACCCTCCATAAATCCCCGCGTCTGGTAATTACTTTACGTATACGTGAA 59705
Qy       4321 ACTTTTGTACTTCTGTCTGAGAAATCCAACAGAGGTGGCCGCGTATAAGCAGTTCCGCTT  4380
Db      59706 ACTTTTGTACTTCTGTCTGAGAAATCCAACAGAGGTGGCCGCGTATAAGCAGTTCCGCTT 59765
Qy       4381 TCACTACCGTTTATGTGAAAGTCTTCAAACTCGGCTAAAAACTCATCAAGACTATATTGC  4440
Db      59766 TCACTACCGTTTATGTGAAAGTCTTCAAACTCGGCTAAAAACTCATCAAGACTATATTGC 59825
Qy       4441 ATATAATTCAACAATTGAAGCTCTTCTCTGGGGACACTGTAATTTAACTTGTACAAAATC  4500
Db      59826 ATATAATTCAACAATTGAAGCTCTTCTCTGGGGACACTGTAATTTAACTTGTACAAAATC 59885
```

```
Qy      4501 CTTATAAAAAACGCCCTAAGGTGAAATCCATTTACACATATTTCTGTGGTAGAGCGATCA 4560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     59886 CTTATAAAAAACGCCCTAAGGTGAAATCCATTTACACATATTTCTGTGGTAGAGCGATCA 59945

Qy      4561 GCTTTATATCGCAATGTTGCTGTACACAGCCCTGGACATAAAGCTAAATTTGTTAAATGA 4620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     59946 GCTTTATATCGCAATGTTGCTGTACACAGCCCTGGACATAAAGCTAAATTTGTTAAATGA 60005

Qy      4621 AAAGAAGCAATCTCAAACGGCTCCGGAAATATAAAGTCACTAAGTGGTATTATTTTACAA 4680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60006 AAAGAAGCAATCTCAAACGGCTCCGGAAATATAAAGTCACTAAGTGGTATTATTTTACAA 60065

Qy      4681 CATGGATTTACTACAAGCCCGGTACAACATGTAATTAAAAAAAATCCACATGTAATAAAT 4740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60066 CATGGATTTACTACAAGCCCGGTACAACATGTAATTAAAAAAAATCCACATGTAATAAAT 60125

Qy      4741 AAACTTAAGGTATACATACTTCTCATGTTTTATACATAATATGAACGCTAATGCTGTACT 4800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60126 AAACTTAAGGTATACATACTTCTCATGTTTTATACATAATATGAACGCTAATGCTGTACT 60185

Qy      4801 TATATAATATGTATAGAATTATGCCAATACAGTCATGGTAAGTATATTTTAAGGCATAAT 4860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60186 TATATAATATGTATAGAATTATGCCAATACAGTCATGGTAAGTATATTTTAAGGCATAAT 60245

Qy      4861 GCAAACATCAAATAATATAACACAAAATGCACGCTTCCGGATATCCGCCCAGCTGTTAAC 4920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60246 GCAAACATCAAATAATATAACACAAAATGCACGCTTCCGGATATCCGCCCAGCTGTTAAC 60305

Qy      4921 CCGAAAATACAGAAATGACTACACAAACACACCTGAAACCAATTTTATTCTCAACATATG 4980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60306 CCGAAAATACAGAAATGACTACACAAACACACCTGAAACCAATTTTATTCTCAACATATG 60365

Qy      4981 CAACAATATTTCAGGGTAACACCATGTTAATAAAATACGCAGCATGCACATTTTTAGCTA 5040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60366 CAACAATATTTCAGGGTAACACCATGTTAATAAAATACGCAGCATGCACATTTTTAGCTA 60425

Qy      5041 AGCTCCTCAAAGACAATTTCCTCTTCAATTGATGACTCATCAACCTCGGTTTCCGTGTTA 5100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60426 AGCTCCTCAAAGACAATTTCCTCTTCAATTGATGACTCATCAACCTCGGTTTCCGTGTTA 60485

Qy      5101 ATGTCGGAAATATATGAGTCCAGGATACCCTCGTCATCGCCACAAATTAATTCCAGGGCT 5160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60486 ATGTCGGAAATATATGAGTCCAGGATACCCTCGTCATCGCCACAAATTAATTCCAGGGCT 60545

Qy      5161 TGTAAAATATCATCCAGTGAATCAGCTGCTAATGTCAAGGATGTTGTTTCTAATTTTAGC 5220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60546 TGTAAAATATCATCCAGTGAATCAGCTGCTAATGTCAAGGATGTTGTTTCTAATTTTAGC 60605

Qy      5221 CCATCCAAATGGTGAGTGGGTGACACTTTGGGACAAGACCCCGTAAAATTGGCACGCTCC 5280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60606 CCATCCAAATGGTGAGTGGGTGACACTTTGGGACAAGACCCCGTAAAATTGGCACGCTCC 60665

Qy      5281 ACACAAACGGAAGGATGCATTTGATCGTTCACGAGGGGCAGCATATTTATTTTTATAACA 5340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60666 ACACAAACGGAAGGATGCATTTGATCGTTCACGAGGGGCAGCATATTTATTTTTATAACA 60725

Qy      5341 TCAGGAACCGGACCCGTCAATGACGCACATATGTCCATTAAGATTGGAGTCGTGACACTG 5400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60726 TCAGGAACCGGACCCGTCAATGACGCACATATGTCCATTAAGATTGGAGTCGTGACACTG 60785

Qy      5401 GTTGAATTGACGGTGAAAAAAAAATGTATTCGTGTCACATGCGTAACTATTTTTTAACCAC 5460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60786 GTTGAATTGACGGTGAAAAAAAAATGTATTCGTGTCACATGCGTAACTATTTTTTAACCAC 60845

Qy      5461 ACGAGGGCGGAATAAAACGGGTCCCCGGCGCTAAGGCCGTGAATGCGGCAGGCACGGCTC 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60846 ACGAGGGCGGAATAAAACGGGTCCCCGGCGCTAAGGCCGTGAATGCGGCAGGCACGGCTC 60905

Qy      5521 ATTGCCATTTCCAGATTTGTTAAATCTGTTCGAACATACGAGTCTAAAATCAAGTATAAC 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60906 ATTGCCATTTCCAGATTTGTTAAATCTGTTCGAACATACGAGTCTAAAATCAAGTATAAC 60965

Qy      5581 CACGACAAGTTCAAACAGGAGGCAACGCGGCTGGTTTCCACCCGGACATCCCCGCTCACA 5640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60966 CACGACAAGTTCAAACAGGAGGCAACGCGGCTGGTTTCCACCCGGACATCCCCGCTCACA 61025

Qy      5641 AACCCGAACTCTCTAGGAACGCACATCAAAGCGTTCAGACAAAATTCCGAAACCGAAACC 5700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61026 AACCCGAACTCTCTAGGAACGCACATCAAAGCGTTCAGACAAAATTCCGAAACCGAAACC 61085

Qy      5701 CTAGTTCTGAGCACCATCTGGTTATAAACTGCGGTTAAGAGACGAGCGGCCACACACTGC 5760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61086 CTAGTTCTGAGCACCATCTGGTTATAAACTGCGGTTAAGAGACGAGCGGCCACACACTGC 61145

Qy      5761 TGCTTGTGAGGATTCAGCTTAGGTGGCCTGCAGGCTTGCTGACAGGCCCGTAGCTGCCTG 5820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61146 TGCTTGTGAGGATTCAGCTTAGGTGGCCTGCAGGCTTGCTGACAGGCCCGTAGCTGCCTG 61205

Qy      5821 GCGGCGCTCGCGCACCTTTGACAGCCGCACGCCAGTTCCAGAGCGAGGCAGAAGCGCTCT 5880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61206 GCGGCGCTCGCGCACCTTTGACAGCCGCACGCCAGTTCCAGAGCGAGGCAGAAGCGCTCT 61265

Qy      5881 TTGCAGTCGCGCCATATCTCGGCCATAGGAGATTCGCTCGCGTGATCTTTTAGGTGCATA 5940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61266 TTGCAGTCGCGCCATATCTCGGCCATAGGAGATTCGCTCGCGTGATCTTTTAGGTGCATA 61325

Qy      5941 TATATCTGTTGGGTAACAAACCTCACTCCCTGTAAAAGGGGAATAAGGTCCGTTCTTTCT 6000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61326 TATATCTGTTGGGTAACAAACCTCACTCCCTGTAAAAGGGGAATAAGGTCCGTTCTTTCT 61385

Qy      6001 ATCGTTTTCTCTGAATCTATAGTTACATCTTTAAGTATAACCAAAGACGCTAAAAACCCA 6060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61386 ATCGTTTTCTCTGAATCTATAGTTACATCTTTAAGTATAACCAAAGACGCTAAAAACCCA 61445

Qy      6061 GGATGGTCGACACCTTTAAAAAATCGCAAAAGCGACCGGATGGCTCTATCCGCGTTTCCA 6120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61446 GGATGGTCGACACCTTTAAAAAATCGCAAAAGCGACCGGATGGCTCTATCCGCGTTTCCA 61505

Qy      6121 TGTCGTTCAAAGGTAGTCACGATGGATCTCCAGTTAGACTCTGTCTCCCTGTTAACACCC 6180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61506 TGTCGTTCAAAGGTAGTCACGATGGATCTCCAGTTAGACTCTGTCTCCCTGTTAACACCC 61565

Qy      6181 TTAACTGGAATGGAAACAGCCATGACACCGTGAACTTCCTGATGTCTCCTAAAACTAACC 6240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61566 TTAACTGGAATGGAAACAGCCATGACACCGTGAACTTCCTGATGTCTCCTAAAACTAACC 61625

Qy      6241 CCGAAACAGAGCTAAATACCAATGACTGTCACCCCTACCCCAAGCCACGCCCCCGTACTA 6300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61626 CCGAAACAGAGCTAAATACCAATGACTGTCACCCCTACCCCAAGCCACGCCCCCGTACTA 61685

Qy      6301 TTAGACCAGGGTGAGTAACCACGCTATCCTTTAAAAACCCATACGTGGAGTTTGTAAGGT 6360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61686 TTAGACCAGGGTGAGTAACCACGCTATCCTTTAAAAACCCATACGTGGAGTTTGTAAGGT 61745

Qy      6361 AAACAGCTCGTGTATATTTCAGACGCCTGTCATAAAATGGATACCGACGACAATCAGGTA 6420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61746 AAACAGCTCGTGTATATTTCAGACGCCTGTCATAAAATGGATACCGACGACAATCAGGTA 61805

Qy      6421 ATTAAACTTTTTTTATTCAAGATTCAGGAGGGCGTGTTTACAGTGGTGTAGGTGGGAGCA 6480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61806 ATTAAACTTTTTTTATTCAAGATTCAGGAGGGCGTGTTTACAGTGGTGTAGGTGGGAGCA 61865

Qy      6481 TATCTCGCCAATGGGAATGGCTGATGAATCCACACTTAGTGCTCGGATCAGGTGCTCTGT 6540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61866 TATCTCGCCAATGGGAATGGCTGATGAATCCACACTTAGTGCTCGGATCAGGTGCTCTGT 61925

Qy      6541 GTACTTTACTAGTGCCGATGAACCTAAATTTGGCGTGTTTAGGTCGTACCAACCAAAAAA 6600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61926 GTACTTTACTAGTGCCGATGAACCTAAATTTGGCGTGTTTAGGTCGTACCAACCAAAAAA 61985

Qy      6601 CTGCCTGGCCTCGCCTTTTAGAAGCTCTATGCAAGCTTTAATCGTGTCCAATAGCTTGTC 6660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61986 CTGCCTGGCCTCGCCTTTTAGAAGCTCTATGCAAGCTTTAATCGTGTCCAATAGCTTGTC 62045

Qy      6661 TTTAAAAATGCAACTCTGGTACATCTTTATGACAGTGGTCCAAAAAAAACAAAGATTTAA 6720
```

```
Db      62046 TTTAAAAATGCAACTCTGGTACATCTTTATGACAGTGGTCCAAAAAAAACAAAGATTTAA 62105
Qy       6721 AAACACGTTAAACTGCGTGTCCTGGTAGTCCTCGTATATAACCTCCTCAACAAGAAAAAA 6780
Db      62106 AAACACGTTAAACTGCGTGTCCTGGTAGTCCTCGTATATAACCTCCTCAACAAGAAAAAA 62165
Qy       6781 TTTTTTAACCAAACTCGCCAGGTACTGAAACGATGGTACGGACAGGTCGTGAAAAGTGTC 6840
Db      62166 TTTTTTAACCAAACTCGCCAGGTACTGAAACGATGGTACGGACAGGTCGTGAAAAGTGTC 62225
Qy       6841 TATCATCTCCTTCACCTCCCCTTTGAAGGTTTTGGTTACACCGACCATGTGAGATAGGCA 6900
Db      62226 TATCATCTCCTTCACCTCCCCTTTGAAGGTTTTGGTTACACCGACCATGTGAGATAGGCA 62285
Qy       6901 CCAGCTCAGGGGGGAGGTCGGATCGTGAGGGGGTAACAATTCGTGGGCGTGGGGGTACTC 6960
Db      62286 CCAGCTCAGGGGGGAGGTCGGATCGTGAGGGGGTAACAATTCGTGGGCGTGGGGGTACTC 62345
Qy       6961 CAGTTTTAGGATCTCCGGCAGGACGCGTACGAGTTCTTCGTCAAATCCTACGCAGCCCGC 7020
Db      62346 CAGTTTTAGGATCTCCGGCAGGACGCGTACGAGTTCTTCGTCAAATCCTACGCAGCCCGC 62405
Qy       7021 ATCCACCAGGGGCAGTACTGAGTTAATAAGTCCCCGAATTCTATCGTTCGTAATTTTGTG 7080
Db      62406 ATCCACCAGGGGCAGTACTGAGTTAATAAGTCCCCGAATTCTATCGTTCGTAATTTTGTG 62465
Qy       7081 GAGCTCCTTTAGACAGTAAACCATGTTGCTCCGACACTGCGGTACGATGAACCGCTCTCC 7140
Db      62466 GAGCTCCTTTAGACAGTAAACCATGTTGCTCCGACACTGCGGTACGATGAACCGCTCTCC 62525
Qy       7141 TTGTCTGTGTTGCATGAGCACCTGGAGTATCTTAGACAGGTATAGGGCCCGTTCCACTCT 7200
Db      62526 TTGTCTGTGTTGCATGAGCACCTGGAGTATCTTAGACAGGTATAGGGCCCGTTCCACTCT 62585
Qy       7201 AAGGGTTGCCGCTTGGTAGAGCGGATTGCGGGCATTGACGGTCAGTGTGTTCATTTCACC 7260
Db      62586 AAGGGTTGCCGCTTGGTAGAGCGGATTGCGGGCATTGACGGTCAGTGTGTTCATTTCACC 62645
Qy       7261 GACTATCATGCCGAGTGGCGGATAGTGGAACTTGTACATATGATTTAAAAGATGGTCTTT 7320
Db      62646 GACTATCATGCCGAGTGGCGGATAGTGGAACTTGTACATATGATTTAAAAGATGGTCTTT 62705
Qy       7321 CCCGTAATGGCGACTCATGGTCGTGTTTTAACGTTCACAGGGTGTTAAAAAACTCAGAAG 7380
Db      62706 CCCGTAATGGCGACTCATGGTCGTGTTTTAACGTTCACAGGGTGTTAAAAAACTCAGAAG 62765
Qy       7381 GTCCTGCGTGGACAACTTTTGCGGACTCTCCGAACCGCTCAGGGATAAGCTGTACCACGC 7440
Db      62766 GTCCTGCGTGGACAACTTTTGCGGACTCTCCGAACCGCTCAGGGATAAGCTGTACCACGC 62825
Qy       7441 GGTGCTTCTTATAAACGAAGCATACGTCAAAAAACATGACCCACGTGACCTAAATAAGCT 7500
Db      62826 GGTGCTTCTTATAAACGAAGCATACGTCAAAAAACATGACCCACGTGACCTAAATAAGCT 62885
Qy       7501 CACAGAAGAAGTCTGTCGATGTATTGTAATGGAATGTGCCTCCCTGGGTCCCATCTCCGG 7560
Db      62886 CACAGAAGAAGTCTGTCGATGTATTGTAATGGAATGTGCCTCCCTGGGTCCCATCTCCGG 62945
Qy       7561 GCTCATCGCGGATCTGAATCTCTTTAACCTGTTCTGCCTCTACCGGGGTTCCCGGGTTAA 7620
Db      62946 GCTCATCGCGGATCTGAATCTCTTTAACCTGTTCTGCCTCTACCGGGGTTCCCGGGTTAA 63005
Qy       7621 AACCCGCGGGGCCGCCACCTGTAACGTCCCGTGCGCAGAATGCGCCCAGGGCATCGTGAG 7680
Db      63006 AACCCGCGGGGCCGCCACCTGTAACGTCCCGTGCGCAGAATGCGCCCAGGGCATCGTGAG 63065
Qy       7681 AATTCTCACCGAGAGGGCCCTTTGTTGTACCGAAAAGATGTTCATAGCGTCTGCCTGCAG 7740
Db      63066 AATTCTCACCGAGAGGGCCCTTTGTTGTACCGAAAAGATGTTCATAGCGTCTGCCTGCAG 63125
Qy       7741 CGGGGTCGTTATACCGCCCCAGTTGGCCAGAGTTCTGCACGACGTCTACGCGGAAATGAA 7800
Db      63126 CGGGGTCGTTATACCGCCCCAGTTGGCCAGAGTTCTGCACGACGTCTACGCGGAAATGAA 63185
Qy       7801 GGCCAAGTGTTTGGGGGCGTGGCGTCGTCTCATATGTTGCAGGCGGCCTATCATGGCTAT 7860
Db      63186 GGCCAAGTGTTTGGGGGCGTGGCGTCGTCTCATATGTTGCAGGCGGCCTATCATGGCTAT 63245
Qy       7861 TGCCGACTCGGTCCTCGTCACTTATAATACCCTGGATGCCGAGGGAAAACTGGAACTTAG 7920
Db      63246 TGCCGACTCGGTCCTCGTCACTTATAATACCCTGGATGCCGAGGGAAAACTGGAACTTAG 63305
Qy       7921 ACTCAAAGCACTGTGCAAACTTGTTTTTCAACCCATCTTTCTTCAAAGAATCTTAGCCCC 7980
Db      63306 ACTCAAAGCACTGTGCAAACTTGTTTTTCAACCCATCTTTCTTCAAAGAATCTTAGCCCC 63365
Qy       7981 TATGCAGCTACTGGCCAACGGGAAGATGGTTCCTGACAATTATTTTACCATCACCGGTAC 8040
Db      63366 TATGCAGCTACTGGCCAACGGGAAGATGGTTCCTGACAATTATTTTACCATCACCGGTAC 63425
Qy       8041 GGCCGAGAAGAGGCGCCCTGTCGTGACTGGAAGTACTAGCGGGATGACGTGTCCGGGAAG 8100
Db      63426 GGCCGAGAAGAGGCGCCCTGTCGTGACTGGAAGTACTAGCGGGATGACGTGTCCGGGAAG 63485
Qy       8101 CAGCCTTGTCCCCGATTCCTTAATCCTGCCAGTATGCGAACCGGGGCTTCTCCCGGCACC 8160
Db      63486 CAGCCTTGTCCCCGATTCCTTAATCCTGCCAGTATGCGAACCGGGGCTTCTCCCGGCACC 63545
Qy       8161 CCTGGTTGACCTCAGTAATGTCTTAGAAAATCCAGAAATCATCCTCAGCGCCCCACCCCT 8220
Db      63546 CCTGGTTGACCTCAGTAATGTCTTAGAAAATCCAGAAATCATCCTCAGCGCCCCACCCCT 63605
Qy       8221 GAGTCAATTTGTCATCACAAACACGCACCCCAGTCTGCCTCAGTCAGTCAGCATTATTAC 8280
Db      63606 GAGTCAATTTGTCATCACAAACACGCACCCCAGTCTGCCTCAGTCAGTCAGCATTATTAC 63665
Qy       8281 GCCAACCCAGGGCGTTGTTCCCGGCCAATGTTTTATGGACACGTGGAAAGCGGTGTCACA 8340
Db      63666 GCCAACCCAGGGCGTTGTTCCCGGCCAATGTTTTATGGACACGTGGAAAGCGGTGTCACA 63725
Qy       8341 GAGCATTCACCACCAGGCACAGACGCCTATTTTGGCCGCCGCACTAACCGGTTCGACATC 8400
Db      63726 GAGCATTCACCACCAGGCACAGACGCCTATTTTGGCCGCCGCACTAACCGGTTCGACATC 63785
Qy       8401 TGCGGCCCCTGGCCCGCATATCGCATGTTCCCCAGTTGCCGGCACGTCTCGGCAGGTGGA 8460
Db      63786 TGCGGCCCCTGGCCCGCATATCGCATGTTCCCCAGTTGCCGGCACGTCTCGGCAGGTGGA 63845
Qy       8461 AGGGTCCGCGGGCGTCGATTGCGGGAAACCAGCATGCGTTCCGCAGCCCGCGTTACCGCC 8520
Db      63846 AGGGTCCGCGGGCGTCGATTGCGGGAAACCAGCATGCGTTCCGCAGCCCGCGTTACCGCC 63905
Qy       8521 CAATGTCCCCGCCAAGAGGATGGAAACGGTAGCACAGTTGGGAAACGCTCCCGTAAAAAA 8580
Db      63906 CAATGTCCCCGCCAAGAGGATGGAAACGGTAGCACAGTTGGGAAACGCTCCCGTAAAAAA 63965
Qy       8581 CGTCCACATCGGAGGCCGCGTATACGCTCCACTGGTTAATATACCAATAATAGACTTAAC 8640
Db      63966 CGTCCACATCGGAGGCCGCGTATACGCTCCACTGGTTAATATACCAATAATAGACTTAAC 64025
Qy       8641 GTCCCCGTCAGGGTCCGGCCAGAGTCCGGCCGATATCGCCAACACTCCAGAGTCCCGCAT 8700
Db      64026 GTCCCCGTCAGGGTCCGGCCAGAGTCCGGCCGATATCGCCAACACTCCAGAGTCCCGCAT 64085
Qy       8701 GGCGGCCGGCTCTCCGCCCTTCGCCGAAACCGCCGCAACGGTCCCCGCTAAGAGAAAGCA 8760
Db      64086 GGCGGCCGGCTCTCCGCCCTTCGCCGAAACCGCCGCAACGGTCCCCGCTAAGAGAAAGCA 64145
Qy       8761 GCCACGCGAGGACGTGGCAGACAAAAGACTGAAGGGCGACGTTCGGGGCGCCGCAACAGT 8820
Db      64146 GCCACGCGAGGACGTGGCAGACAAAAGACTGAAGGGCGACGTTCGGGGCGCCGCAACAGT 64205
Qy       8821 AAACCACCCTTTCCCGGGACCGTCCGGGATGCGCGTTCGCGAGCAGGGCTTATTCGATTT 8880
Db      64206 AAACCACCCTTTCCCGGGACCGTCCGGGATGCGCGTTCGCGAGCAGGGCTTATTCGATTT 64265
```

```
Qy       8881 AATCGAAAGCTCCACGGATGTAACCGCGAACGCATCTGGACCCAAAAACGACGACGACAT 8940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64266 AATCGAAAGCTCCACGGATGTAACCGCGAACGCATCTGGACCCAAAAACGACGACGACAT 64325

Qy       8941 GCTAGCGGCTATCCTACAGGACCTGTATGGACTACAGTCCCCCCCGGCCATCGATTCCCC 9000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64326 GCTAGCGGCTATCCTACAGGACCTGTATGGACTACAGTCCCCCCCGGCCATCGATTCCCC 64385

Qy       9001 CTCCAGCAACTCGGACAATGAGGAGATATTTCCAGAGGTTAGTCCGCCATCTAGCGGCCA 9060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64386 CTCCAGCAACTCGGACAATGAGGAGATATTTCCAGAGGTTAGTCCGCCATCTAGCGGCCA 64445

Qy       9061 CGGATCGCCTTGA 9073
             |||||||||||||
Db      64446 CGGATCGCCTTGA 64458
```

RESULT 2
AAC64754
ID AAC64754 standard; DNA; 133719 BP.
XX
AC AAC64754;
XX
DT 28-FEB-2001 (first entry)
XX
DE Macaca mulatta rhadinovirus 17577 (RRV) genome sequence SEQ ID NO:1.
XX
KW Macaca mulatta rhadinovirus 17577; RRV; rhesus macaque rhadinovirus;
KW genome; Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6;
KW IL-6; macrophage inflammatory protein; MIP; diagnosis; vaccine;
KW cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma;
KW lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy;
KW splenomegaly; hypergammaglobinulinaemia; autoimmune haemolytic anaemia;
KW ds.
XX
OS Macaca mulatta rhadinovirus 17577.
XX
PN WO200028040-A2.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26260.
XX
PR 06-NOV-1998; 98US-0107507.
PR 20-NOV-1998; 98US-0109409.
XX
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI Wong SW, Axthelm MK, Searles RP;
XX
DR WPI; 2000-376552/32.
XX
PT New rhesus rhadino virus for producing non-human primate model useful
PT for testing potential treatments and efficacy of the candidate vaccine
PT for conditions associated with RRV infection
XX
PS Claim 2; Page 83-122; 141pp; English.
XX
CC The present invention describes a novel rhesus macaque rhadinovirus
CC called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the
CC RRV genome sequence, and AAB53123 to AAB53204 represent the proteins
CC encoded by the genome sequence. The present invention also specifically
CC claims the individual open reading frame (ORF) nucleotide sequences from
CC the genome which encode the individual proteins, but these sequences are
CC not given. A non-human animal infected with RRV can be used for testing
CC the efficacy of drug in the treatment of condition associated with
CC infection with RRV such as Kaposi's sarcoma, lymphoproliferative
CC disorders, B-cell hyperplasia, lymphadenopathy, splenomegaly,
CC hypergammaglobinulinaemia or autoimmune haemolytic anaemia, by
CC administering the drug to a immuno-compromised non-human primate
CC preferably Rhesus macaque monkey obtained by as a result of infection
CC by Simian Immunodeficiency Virus (SIV). RRV is useful for producing
CC non-human primate model for testing potential treatments for conditions
CC associated with RRV infection. It is also useful for testing the
CC efficacy of the candidate vaccine against RRV infection or conditions
CC associated with its infection by administering the vaccine to the
CC subject capable of infection with RRV, inoculating the subject with RRV
CC and observing the effect of vaccine. AAC64755 to AAC64765 and AAB53205
CC to AAB53213 represent sequence used in the exemplification of the
CC present invention.
XX
SQ Sequence 133719 BP; 32746 A; 35648 C; 34521 G; 30804 T; 0 other;

Query Match 100.0%; Score 9073; DB 21; Length 133719;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 9073; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGGAGAGTTCCGTCGGATGGACCAAACACGTCGAACCAAATCCGGGGTTCATCTTGAAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60966 ATGGAGAGTTCCGTCGGATGGACCAAACACGTCGAACCAAATCCGGGGTTCATCTTGAAC 61025

Qy         61 ATGACGTCCGATGCCAAAGTCAGGGGTGTCGTGGATCACGTCAGTCGCCTGTCAAATATA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61026 ATGACGTCCGATGCCAAAGTCAGGGGTGTCGTGGATCACGTCAGTCGCCTGTCAAATATA 61085

Qy        121 ACTACCAGCCCACCGGAAATGGGTTGGTACGACCTGGCCTTCGATCCGGCTGAAGACTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61086 ACTACCAGCCCACCGGAAATGGGTTGGTACGACCTGGCCTTCGATCCGGCTGAAGACTCC 61145

Qy        181 GGGCCGTTCTTGCCGTTTACCGTTTATCTAATTACGGGAACTGCTGGTGCTGGGAAAAGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61146 GGGCCGTTCTTGCCGTTTACCGTTTATCTAATTACGGGAACTGCTGGTGCTGGGAAAAGT 61205

Qy        241 ACCAGCATATCGGCCCTGTACCAAAATTTAAACTGCCTGATCACGGGCGCGACCACCATA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61206 ACCAGCATATCGGCCCTGTACCAAAATTTAAACTGCCTGATCACGGGCGCGACCACCATA 61265

Qy        301 GCCGCACAGAACCTATCGCGTCGCCTAAAGACGTTCTGTCCCACGATCTTCAGCGCTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61266 GCCGCACAGAACCTATCGCGTCGCCTAAAGACGTTCTGTCCCACGATCTTCAGCGCTTTT 61325

Qy        361 GGCTTTAAGAGCCGACACATCAATATAGCCGTCAGAAAAGCTCATCAGACCGGAGCCGTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61326 GGCTTTAAGAGCCGACACATCAATATAGCCGTCAGAAAAGCTCATCAGACCGGAGCCGTA 61385

Qy        421 TCCATAGAGCAAATTCAGCAACAGGAGCTATCGAAGTATTGGCCGGTTATAGTGGACATT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61386 TCCATAGAGCAAATTCAGCAACAGGAGCTATCGAAGTATTGGCCGGTTATAGTGGACATT 61445

Qy        481 ATGAAAGAGGTTATGGCGAAAAAACCCAATGGCATGTACGGGACTATATCCAACGCGAAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61446 ATGAAAGAGGTTATGGCGAAAAAACCCAATGGCATGTACGGGACTATATCCAACGCGAAT 61505

Qy        541 TTTGAAACCCTCTCGAGAATGACCGGACCGTGTTTATGGACTTCCAATATTATTGTAATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61506 TTTGAAACCCTCTCGAGAATGACCGGACCGTGTTTATGGACTTCCAATATTATTGTAATC 61565

Qy        601 GACGAGGCCGGAACCCTGTCCTCTTACATACTCACCACCGTCGTGTTCTTTTACTGGTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61566 GACGAGGCCGGAACCCTGTCCTCTTACATACTCACCACCGTCGTGTTCTTTTACTGGTTC 61625

Qy        661 CTAAACAGCTGGCTAAATACCCCTCTTTACCGCCAGGGGGCGGTTCCGTGCATAGTATGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61626 CTAAACAGCTGGCTAAATACCCCTCTTTACCGCCAGGGGGCGGTTCCGTGCATAGTATGC 61685

Qy        721 GTCGGTTCGCCAACGCAGACAAACGCGTTCCAGTCAACTTACAACCACGGGACGCAAAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61686 GTCGGTTCGCCAACGCAGACAAACGCGTTCCAGTCAACTTACAACCACGGGACGCAAAAG 61745

Qy        781 ACGGAGATATCGTCGTGCGAGAACATCCTAACATTCATGATAGGGAAGAAGGTCGTATCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61746 ACGGAGATATCGTCGTGCGAGAACATCCTAACATTCATGATAGGGAAGAAGGTCGTATCT 61805

Qy        841 GAGTACGTACACTTGGAGAGAAACTGGGCGCTGTTTATAAACAACAAGCGCTGCACCGAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      61806 GAGTACGTACACTTGGAGAGAAACTGGGCGCTGTTTATAAACAACAAGCGCTGCACCGAT 61865

Qy        901 CTGCAGTTCGGACACCTTCTAAAAATTTTAGAGTATAATCTTCCCATTCCTGACGAAGTC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61866 CTGCAGTTCGGACACCTTCTAAAAATTTTAGAGTATAATCTTCCCATTCCTGACGAAGTC 61925

Qy        961 ATGAGTTACGTAGACAGATTTGTCGTTCCTAAAAGTAAGATTATGGATCCTTTAGAATAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61926 ATGAGTTACGTAGACAGATTTGTCGTTCCTAAAAGTAAGATTATGGATCCTTTAGAATAC 61985

Qy       1021 ATTGGCTGGACCCGACTCTTTTTGTCACATAGCGAGGTAAAGGCGTATCTAACAAACCTA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61986 ATTGGCTGGACCCGACTCTTTTTGTCACATAGCGAGGTAAAGGCGTATCTAACAAACCTA 62045

Qy       1081 CACACATGTCTAACGCTAGGGGGCGATACCAGGGACACGAAGCTCTTTACCTGTCCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62046 CACACATGTCTAACGCTAGGGGGCGATACCAGGGACACGAAGCTCTTTACCTGTCCCGTG 62105

Qy       1141 GTGTGCGAGGTGTTTGTGAAGCCGTTTGAGGAATACAAGCGGGCCGTCAACCTCACCAAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62106 GTGTGCGAGGTGTTTGTGAAGCCGTTTGAGGAATACAAGCGGGCCGTCAACCTCACCAAC 62165

Qy       1201 CTCACCGTGACCGAGTGGGTGACAAAAAACCTCTTTAAGTTAAGTAATTATTCGCAGTTT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62166 CTCACCGTGACCGAGTGGGTGACAAAAAACCTCTTTAAGTTAAGTAATTATTCGCAGTTT 62225

Qy       1261 GTGGACCAGGACATGTCCATAGTCGCCACGGAATCCACTGAACGTTCGACCCAGGTTACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62226 GTGGACCAGGACATGTCCATAGTCGCCACGGAATCCACTGAACGTTCGACCCAGGTTACC 62285

Qy       1321 TTTATCACCAAGTTTGTTAAAAACAGCCACGTGTCTTTAAATGGAAAAACAAAAAAATGC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62286 TTTATCACCAAGTTTGTTAAAAACAGCCACGTGTCTTTAAATGGAAAAACAAAAAAATGC 62345

Qy       1381 ATATGTGGGTTTCAGGGTACGTACTTTGAGTTCAAAAGAATCCTAGACAGCGAACTCTTC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62346 ATATGTGGGTTTCAGGGTACGTACTTTGAGTTCAAAAGAATCCTAGACAGCGAACTCTTC 62405

Qy       1441 GTGGAGACGCATTCGCAAGACCGTCCGGAATACGTATACGGTTTCTTAAACACGCTACTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62406 GTGGAGACGCATTCGCAAGACCGTCCGGAATACGTATACGGTTTCTTAAACACGCTACTG 62465

Qy       1501 TACAACGCCATGTACTCGTTTCACGCGTACGGCGTGACTAGGTCACACGAGAAATATCTG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62466 TACAACGCCATGTACTCGTTTCACGCGTACGGCGTGACTAGGTCACACGAGAAATATCTG 62525

Qy       1561 CAAGACCTAAAATTTGCACCCCTCCCGGCCGCTCTGGCAACCGGGCGTGTAGACCTTCAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62526 CAAGACCTAAAATTTGCACCCCTCCCGGCCGCTCTGGCAACCGGGCGTGTAGACCTTCAA 62585

Qy       1621 ACGGTTCGTGAAGAGTTAAACCTGGAAGACGACATCTTCTACCACGTGTGCAGTCCCCCG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62586 ACGGTTCGTGAAGAGTTAAACCTGGAAGACGACATCTTCTACCACGTGTGCAGTCCCCCG 62645

Qy       1681 CCACCCGCGGGTATCACCTCCCTCCAGGTTTTGGTCGACACGTACTGCGCCCTAAAGGAC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62646 CCACCCGCGGGTATCACCTCCCTCCAGGTTTTGGTCGACACGTACTGCGCCCTAAAGGAC 62705

Qy       1741 GTGTTCGCCTCCAGAATAAAGGTGGCGTGTCGCTGGTTTGGCGGGGAGTTTGAGAAGGAA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62706 GTGTTCGCCTCCAGAATAAAGGTGGCGTGTCGCTGGTTTGGCGGGGAGTTTGAGAAGGAA 62765

Qy       1801 ACGTTTTCCGCGTTTACGGTTAACATGGTCGTAAGGGACGGAGTTGACTTTGTCTCCCCT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62766 ACGTTTTCCGCGTTTACGGTTAACATGGTCGTAAGGGACGGAGTTGACTTTGTCTCCCCT 62825

Qy       1861 TCAGAACGTCTCAACGGGCTGTTGGCGTTTGCATCGACCGTTGAATCGTATAAAATTAAG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62826 TCAGAACGTCTCAACGGGCTGTTGGCGTTTGCATCGACCGTTGAATCGTATAAAATTAAG 62885

Qy       1921 GGGTACACGTTTTTACCGGTAGCGTTCGGTCGCTGTCAGGGTTTGCCACTCAGTGATGAC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62886 GGGTACACGTTTTTACCGGTAGCGTTCGGTCGCTGTCAGGGTTTGCCACTCAGTGATGAC 62945

Qy       1981 CTCAGGAAGAAGATGCCCTCCCTGGTCGTGCAGGACTCTAGCGGTTTTATCGCGTGCCTA 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62946 CTCAGGAAGAAGATGCCCTCCCTGGTCGTGCAGGACTCTAGCGGTTTTATCGCGTGCCTA 63005

Qy       2041 GAGAATAACATAACCAAATTGACCGAAACCATGGAGGACGGGAGCGTTTTCCAAGTGTGC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63006 GAGAATAACATAACCAAATTGACCGAAACCATGGAGGACGGGAGCGTTTTCCAAGTGTGC 63065

Qy       2101 TGTGCGGGGGACTATGGGGTCAGCTCAAATTTAGCCATGACCATCGTAAAGGCACAGGGA 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63066 TGTGCGGGGGACTATGGGGTCAGCTCAAATTTAGCCATGACCATCGTAAAGGCACAGGGA 63125

Qy       2161 ATGTCGTTGGAGCGAGTAGCCGTAGTATTTGGGTCCCACAAGAACGTCCAGACAAGCCAC 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63126 ATGTCGTTGGAGCGAGTAGCCGTAGTATTTGGGTCCCACAAGAACGTCCAGACAAGCCAC 63185

Qy       2221 GTGTATGTAGCAATATCAAGGGCTGTTAACTCAAATTATTTGGTCATGGACAGCAACCCC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63186 GTGTATGTAGCAATATCAAGGGCTGTTAACTCAAATTATTTGGTCATGGACAGCAACCCC 63245

Qy       2281 CTTAAAACCCTCCTCAGAGAACCAGTCGATAACACCTCCGCCAAGCATATAGTCCGCGCC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63246 CTTAAAACCCTCCTCAGAGAACCAGTCGATAACACCTCCGCCAAGCATATAGTCCGCGCC 63305

Qy       2341 CTCCACAACCCAAACACAACCCTCATCTACTAAATAAATAAAAACAAGCAAATGGAAAAC 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63306 CTCCACAACCCAAACACAACCCTCATCTACTAAATAAATAAAAACAAGCAAATGGAAAAC 63365

Qy       2401 ATTGTGTTTTTATTCAGTCCAACCACGGCCACGGATAGTTGTCATTTCCACACACCGGGG 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63366 ATTGTGTTTTTATTCAGTCCAACCACGGCCACGGATAGTTGTCATTTCCACACACCGGGG 63425

Qy       2461 GGGTGGCCGCCATAGTTTGACGACCAGAAACGCCCGGGCCCTCTGGGTTAGAAGTAGCCC 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63426 GGGTGGCCGCCATAGTTTGACGACCAGAAACGCCCGGGCCCTCTGGGTTAGAAGTAGCCC 63485

Qy       2521 CGCCCCCCGTGGTCGAGGTCATACGTCTCTTAGGGCCCTGGGTTGTACCAACATAAATCA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63486 CGCCCCCCGTGGTCGAGGTCATACGTCTCTTAGGGCCCTGGGTTGTACCAACATAAATCA 63545

Qy       2581 CGTCGTCATCACCGTTGCAGTCCCTGCGTGACGCGCGCCTCTTGCAGGCGGCACGAACAC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63546 CGTCGTCATCACCGTTGCAGTCCCTGCGTGACGCGCGCCTCTTGCAGGCGGCACGAACAC 63605

Qy       2641 CGCCGCTTGTAACGTGATCTGCATCGTCGTCGCTGTCTGAAGATGAACACAGGTCAATCA 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63606 CGCCGCTTGTAACGTGATCTGCATCGTCGTCGCTGTCTGAAGATGAACACAGGTCAATCA 63665

Qy       2701 CATCAGTGGCGCCCGTGCCCTCATCGGCGTCTGATTCGTCCCATGTCTCAACAGAGCCTC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63666 CATCAGTGGCGCCCGTGCCCTCATCGGCGTCTGATTCGTCCCATGTCTCAACAGAGCCTC 63725

Qy       2761 GGTCGTCCACTTCGTCGTCATCTGAAATAATAATGGTCTCTTGTCCTGCAGTTGATCCAG 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63726 GGTCGTCCACTTCGTCGTCATCTGAAATAATAATGGTCTCTTGTCCTGCAGTTGATCCAG 63785

Qy       2821 CAACAACAGGTTCTTCTTCATCGGAAGACCATCCGCCCTCTGCCGGTGATCTAGATATAT 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63786 CAACAACAGGTTCTTCTTCATCGGAAGACCATCCGCCCTCTGCCGGTGATCTAGATATAT 63845

Qy       2881 CACCTGACCAGGCGGGCGAGGCCCGTGGCGAAAAGCCGTCAGCTTCTCCGCCACTTTCAT 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63846 CACCTGACCAGGCGGGCGAGGCCCGTGGCGAAAAGCCGTCAGCTTCTCCGCCACTTTCAT 63905

Qy       2941 CAGTGTCCTCGTAGTCATCTATAAAAAAATCGTCTTCGCTGTCTGATTCAGATGACTGGG 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63906 CAGTGTCCTCGTAGTCATCTATAAAAAAATCGTCTTCGCTGTCTGATTCAGATGACTGGG 63965

Qy       3001 GTGCGCGGTGTTGGCCAGTAACCACCGCAACGGGCCGCGGACCACCAAACGCGGTCAGTG 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63966 GTGCGCGGTGTTGGCCAGTAACCACCGCAACGGGCCGCGGACCACCAAACGCGGTCAGTG 64025
```

```
Qy      3061 GCGCGGGCGCGCTCGGTGGTGTGGTGGCACCTGTCGAAGTAATGACGGTAGGTAGCCGGC 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64026 GCGCGGGCGCGCTCGGTGGTGTGGTGGCACCTGTCGAAGTAATGACGGTAGGTAGCCGGC 64085

Qy      3121 CGCTCTGGTTAAAGGACTCACCGTCCCGTGCGTGCGGATCTATTCTAGGTGGGGAGGAGA 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64086 CGCTCTGGTTAAAGGACTCACCGTCCCGTGCGTGCGGATCTATTCTAGGTGGGGAGGAGA 64145

Qy      3181 CCGGCGGTTCGTCTTCTTCATCCACATCACTGTAGTCGCTGTTCTCCGAATGAGAGTCGG 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64146 CCGGCGGTTCGTCTTCTTCATCCACATCACTGTAGTCGCTGTTCTCCGAATGAGAGTCGG 64205

Qy      3241 TTTCGTCTGGCGTTTTTGGAGTAGTGGGAGGGGAATCTTGGTCAAAAACATCCTCAGGTG 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64206 TTTCGTCTGGCGTTTTTGGAGTAGTGGGAGGGGAATCTTGGTCAAAAACATCCTCAGGTG 64265

Qy      3301 CAGTCATGTCAATCACTGGGGGGCCTCCGTACCCGGGCGGTACCGGCAACCCCGTACGGG 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64266 CAGTCATGTCAATCACTGGGGGGCCTCCGTACCCGGGCGGTACCGGCAACCCCGTACGGG 64325

Qy      3361 ACCCGGCGAACAAAAAACGTCTTGTTCTCTTTTTCCTAGGTGCCCCGGGAATCGGCAGCA 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64326 ACCCGGCGAACAAAAAACGTCTTGTTCTCTTTTTCCTAGGTGCCCCGGGAATCGGCAGCA 64385

Qy      3421 TCCTGGGAGTGGCCGGCGGGGTCCTTGACGGGTCGGATAAGAACATAGCCATGGCCGAAA 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64386 TCCTGGGAGTGGCCGGCGGGGTCCTTGACGGGTCGGATAAGAACATAGCCATGGCCGAAA 64445

Qy      3481 CGTTCACCTGTAAAACGCACTGCCTATCCCCGATCAGTTAATATTCCAGTCAACGGCACC 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64446 CGTTCACCTGTAAAACGCACTGCCTATCCCCGATCAGTTAATATTCCAGTCAACGGCACC 64505

Qy      3541 CCTCCGATTCTGAACTAGATAGTCATTCGCAAGTTTAAAATGGTTGCAGCCTAGAAACTG 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64506 CCTCCGATTCTGAACTAGATAGTCATTCGCAAGTTTAAAATGGTTGCAGCCTAGAAACTG 64565

Qy      3601 CGGCCAGGGAGAACCGGTTGCCGCGTGGGCGGTAGCCAGGGGAGACGGATGGCGCGCCTT 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64566 CGGCCAGGGAGAACCGGTTGCCGCGTGGGCGGTAGCCAGGGGAGACGGATGGCGCGCCTT 64625

Qy      3661 GAGGACGAGATGTCTCTGTCCGTTTATCAGCGTCGCCTTGTCTATAGCCTTTCTTCCCCA 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64626 GAGGACGAGATGTCTCTGTCCGTTTATCAGCGTCGCCTTGTCTATAGCCTTTCTTCCCCA 64685

Qy      3721 CAGCATAAAAACGCACCTCTGAAGCTTGGCAGAAAGGCAGCTTATTATGTAGCTGGTGAA 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64686 CAGCATAAAAACGCACCTCTGAAGCTTGGCAGAAAGGCAGCTTATTATGTAGCTGGTGAA 64745

Qy      3781 CCAATCCCAGCCAAGGTTGGCGTGTGACCCCGCCTTCCCTCTTTCCACCGTCAGAATGGT 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64746 CCAATCCCAGCCAAGGTTGGCGTGTGACCCCGCCTTCCCTCTTTCCACCGTCAGAATGGT 64805

Qy      3841 GTTTAAAAGCAGAACTCCCCGTTTTGCCCAACAGTCTAAGCACCCGTGAGAAGGAGCGGT 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64806 GTTTAAAAGCAGAACTCCCCGTTTTGCCCAACAGTCTAAGCACCCGTGAGAAGGAGCGGT 64865

Qy      3901 GAACCCAGGTACAGTATTGGCTATCTCTTTAAAAATATTTTTGAGGCTTGGAGGTATACT 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64866 GAACCCAGGTACAGTATTGGCTATCTCTTTAAAAATATTTTTGAGGCTTGGAGGTATACT 64925

Qy      3961 GTAGTCCGGAGCAACGCTAAAGGCTAGTCCGGTGGCTTGACCGCGATGGTAAGGATCCTG 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64926 GTAGTCCGGAGCAACGCTAAAGGCTAGTCCGGTGGCTTGACCGCGATGGTAAGGATCCTG 64985

Qy      4021 GCCAAGGATCACGACTTTAATATCCTCCGGTTCGCAACAATACGACCACCACATAATTCT 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64986 GCCAAGGATCACGACTTTAATATCCTCCGGTTCGCAACAATACGACCACCACATAATTCT 65045

Qy      4081 ATCTATCGGTGGATAAATCACGGTTACGTTGCTCATATCCATAACGCGCTTCAGCAGAGC 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65046 ATCTATCGGTGGATAAATCACGGTTACGTTGCTCATATCCATAACGCGCTTCAGCAGAGC 65105

Qy      4141 CGCGAGCTTCTGTTTTAAAAATGGAGAAAGATTTAAAAATTCAAGCCATGAGTCGCTAAG 4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65106 CGCGAGCTTCTGTTTTAAAAATGGAGAAAGATTTAAAAATTCAAGCCATGAGTCGCTAAG 65165

Qy      4201 TAGCAACGTTTGTGTGGAGGGTTCTTCTAACACTTCTGGTGACATTTTACTCCAAACTAT 4260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65166 TAGCAACGTTTGTGTGGAGGGTTCTTCTAACACTTCTGGTGACATTTTACTCCAAACTAT 65225

Qy      4261 TGTTTTAAGCCAACCCTCCATAAATCCCCGCGTCTGGTAATTACTTTACGTATACGTGAA 4320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65226 TGTTTTAAGCCAACCCTCCATAAATCCCCGCGTCTGGTAATTACTTTACGTATACGTGAA 65285

Qy      4321 ACTTTTGTACTTCTGTCTGAGAAATCCAACAGAGGTGGCCGCGTATAAGCAGTTCCGCTT 4380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65286 ACTTTTGTACTTCTGTCTGAGAAATCCAACAGAGGTGGCCGCGTATAAGCAGTTCCGCTT 65345

Qy      4381 TCACTACCGTTTATGTGAAAGTCTTCAAACTCGGCTAAAAACTCATCAAGACTATATTGC 4440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65346 TCACTACCGTTTATGTGAAAGTCTTCAAACTCGGCTAAAAACTCATCAAGACTATATTGC 65405

Qy      4441 ATATAATTCAACAATTGAAGCTCTTCTCTGGGGACACTGTAATTTAACTTGTACAAAATC 4500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65406 ATATAATTCAACAATTGAAGCTCTTCTCTGGGGACACTGTAATTTAACTTGTACAAAATC 65465

Qy      4501 CTTATAAAAAACGCCCTAAGGTGAAATCCATTTACACATATTTCTGTGGTAGAGCGATCA 4560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65466 CTTATAAAAAACGCCCTAAGGTGAAATCCATTTACACATATTTCTGTGGTAGAGCGATCA 65525

Qy      4561 GCTTTATATCGCAATGTTGCTGTACACAGCCCTGGACATAAAGCTAAATTTGTTAAATGA 4620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65526 GCTTTATATCGCAATGTTGCTGTACACAGCCCTGGACATAAAGCTAAATTTGTTAAATGA 65585

Qy      4621 AAAGAAGCAATCTCAAACGGCTCCGGAAATATAAAGTCACTAAGTGGTATTATTTTACAA 4680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65586 AAAGAAGCAATCTCAAACGGCTCCGGAAATATAAAGTCACTAAGTGGTATTATTTTACAA 65645

Qy      4681 CATGGATTTACTACAAGCCCGGTACAACATGTAATTAAAAAAAATCCACATGTAATAAAT 4740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65646 CATGGATTTACTACAAGCCCGGTACAACATGTAATTAAAAAAAATCCACATGTAATAAAT 65705

Qy      4741 AAACTTAAGGTATACATACTTCTCATGTTTTATACATAATATGAACGCTAATGCTGTACT 4800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65706 AAACTTAAGGTATACATACTTCTCATGTTTTATACATAATATGAACGCTAATGCTGTACT 65765

Qy      4801 TATATAATATGTATAGAATTATGCCAATACAGTCATGGTAAGTATATTTTAAGGCATAAT 4860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65766 TATATAATATGTATAGAATTATGCCAATACAGTCATGGTAAGTATATTTTAAGGCATAAT 65825

Qy      4861 GCAAACATCAAATAATATAACACAAAATGCACGCTTCCGGATATCCGCCCAGCTGTTAAC 4920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65826 GCAAACATCAAATAATATAACACAAAATGCACGCTTCCGGATATCCGCCCAGCTGTTAAC 65885

Qy      4921 CCGAAAATACAGAAATGACTACACAAACACACCTGAAACCAATTTTATTCTCAACATATG 4980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65886 CCGAAAATACAGAAATGACTACACAAACACACCTGAAACCAATTTTATTCTCAACATATG 65945

Qy      4981 CAACAATATTTCAGGGTAACACCATGTTAATAAAATACGCAGCATGCACATTTTTAGCTA 5040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65946 CAACAATATTTCAGGGTAACACCATGTTAATAAAATACGCAGCATGCACATTTTTAGCTA 66005

Qy      5041 AGCTCCTCAAAGACAATTTCCTCTTCAATTGATGACTCATCAACCTCGGTTTCCGTGTTA 5100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     66006 AGCTCCTCAAAGACAATTTCCTCTTCAATTGATGACTCATCAACCTCGGTTTCCGTGTTA 66065

Qy      5101 ATGTCGGAAATATATGAGTCCAGGATACCCTCGTCATCGCCACAAATTAATTCCAGGGCT 5160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     66066 ATGTCGGAAATATATGAGTCCAGGATACCCTCGTCATCGCCACAAATTAATTCCAGGGCT 66125

Qy      5161 TGTAAAATATCATCCAGTGAATCAGCTGCTAATGTCAAGGATGTTGTTTCTAATTTTAGC 5220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     66126 TGTAAAATATCATCCAGTGAATCAGCTGCTAATGTCAAGGATGTTGTTTCTAATTTTAGC 66185

Qy      5221 CCATCCAAATGGTGAGTGGGTGACACTTTGGGACAAGACCCCGTAAAATTGGCACGCTCC 5280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      66186 CCATCCAAATGGTGAGTGGGTGACACTTTGGGACAAGACCCCGTAAAATTGGCACGCTCC 66245

Qy       5281 ACACAAACGGAAGGATGCATTTGATCGTTCACGAGGGGCAGCATATTTATTTTTATAACA 5340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66246 ACACAAACGGAAGGATGCATTTGATCGTTCACGAGGGGCAGCATATTTATTTTTATAACA 66305

Qy       5341 TCAGGAACCGGACCCGTCAATGACGCACATATGTCCATTAAGATTGGAGTCGTGACACTG 5400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66306 TCAGGAACCGGACCCGTCAATGACGCACATATGTCCATTAAGATTGGAGTCGTGACACTG 66365

Qy       5401 GTTGAATTGACGGTGAAAAAAAATGTATTCGTGTCACATGCGTAACTATTTTTTAACCAC 5460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66366 GTTGAATTGACGGTGAAAAAAAATGTATTCGTGTCACATGCGTAACTATTTTTTAACCAC 66425

Qy       5461 ACGAGGGCGGAATAAAACGGGTCCCCGGCGCTAAGGCCGTGAATGCGGCAGGCACGGCTC 5520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66426 ACGAGGGCGGAATAAAACGGGTCCCCGGCGCTAAGGCCGTGAATGCGGCAGGCACGGCTC 66485

Qy       5521 ATTGCCATTTCCAGATTTGTTAAATCTGTTCGAACATACGAGTCTAAAATCAAGTATAAC 5580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66486 ATTGCCATTTCCAGATTTGTTAAATCTGTTCGAACATACGAGTCTAAAATCAAGTATAAC 66545

Qy       5581 CACGACAAGTTCAAACAGGAGGCAACGCGGCTGGTTTCCACCCGGACATCCCCGCTCACA 5640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66546 CACGACAAGTTCAAACAGGAGGCAACGCGGCTGGTTTCCACCCGGACATCCCCGCTCACA 66605

Qy       5641 AACCCGAACTCTCTAGGAACGCACATCAAAGCGTTCAGACAAAATTCCGAAACCGAAACC 5700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66606 AACCCGAACTCTCTAGGAACGCACATCAAAGCGTTCAGACAAAATTCCGAAACCGAAACC 66665

Qy       5701 CTAGTTCTGAGCACCATCTGGTTATAAACTGCGGTTAAGAGACGAGCGGCCACACACTGC 5760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66666 CTAGTTCTGAGCACCATCTGGTTATAAACTGCGGTTAAGAGACGAGCGGCCACACACTGC 66725

Qy       5761 TGCTTGTGAGGATTCAGCTTAGGTGGCCTGCAGGCTTGCTGACAGGCCCGTAGCTGCCTG 5820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66726 TGCTTGTGAGGATTCAGCTTAGGTGGCCTGCAGGCTTGCTGACAGGCCCGTAGCTGCCTG 66785

Qy       5821 GCGGCGCTCGCGCACCTTTGACAGCCGCACGCCAGTTCCAGAGCGAGGCAGAAGCGCTCT 5880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66786 GCGGCGCTCGCGCACCTTTGACAGCCGCACGCCAGTTCCAGAGCGAGGCAGAAGCGCTCT 66845

Qy       5881 TTGCAGTCGCGCCATATCTCGGCCATAGGAGATTCGCTCGCGTGATCTTTTAGGTGCATA 5940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66846 TTGCAGTCGCGCCATATCTCGGCCATAGGAGATTCGCTCGCGTGATCTTTTAGGTGCATA 66905

Qy       5941 TATATCTGTTGGGTAACAAACCTCACTCCCTGTAAAAGGGGAATAAGGTCCGTTCTTTCT 6000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66906 TATATCTGTTGGGTAACAAACCTCACTCCCTGTAAAAGGGGAATAAGGTCCGTTCTTTCT 66965

Qy       6001 ATCGTTTTCTCTGAATCTATAGTTACATCTTTAAGTATAACCAAAGACGCTAAAAACCCA 6060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66966 ATCGTTTTCTCTGAATCTATAGTTACATCTTTAAGTATAACCAAAGACGCTAAAAACCCA 67025

Qy       6061 GGATGGTCGACACCTTTAAAAAATCGCAAAAGCGACCGGATGGCTCTATCCGCGTTTCCA 6120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67026 GGATGGTCGACACCTTTAAAAAATCGCAAAAGCGACCGGATGGCTCTATCCGCGTTTCCA 67085

Qy       6121 TGTCGTTCAAAGGTAGTCACGATGGATCTCCAGTTAGACTCTGTCTCCCTGTTAACACCC 6180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67086 TGTCGTTCAAAGGTAGTCACGATGGATCTCCAGTTAGACTCTGTCTCCCTGTTAACACCC 67145

Qy       6181 TTAACTGGAATGGAAACAGCCATGACACCGTGAACTTCCTGATGTCTCCTAAAACTAACC 6240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67146 TTAACTGGAATGGAAACAGCCATGACACCGTGAACTTCCTGATGTCTCCTAAAACTAACC 67205

Qy       6241 CCGAAACAGAGCTAAATACCAATGACTGTCACCCCTACCCCAAGCCACGCCCCCCGTACTA 6300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67206 CCGAAACAGAGCTAAATACCAATGACTGTCACCCCTACCCCAAGCCACGCCCCCCGTACTA 67265

Qy       6301 TTAGACCAGGGTGAGTAACCACGCTATCCTTTAAAAACCCATACGTGGAGTTTGTAAGGT 6360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67266 TTAGACCAGGGTGAGTAACCACGCTATCCTTTAAAAACCCATACGTGGAGTTTGTAAGGT 67325

Qy       6361 AAACAGCTCGTGTATATTTCAGACGCCTGTCATAAAATGGATACCGACGACAATCAGGTA 6420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67326 AAACAGCTCGTGTATATTTCAGACGCCTGTCATAAAATGGATACCGACGACAATCAGGTA 67385

Qy       6421 ATTAAACTTTTTTTATTCAAGATTCAGGAGGGCGTGTTTACAGTGGTGTAGGTGGGAGCA 6480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67386 ATTAAACTTTTTTTATTCAAGATTCAGGAGGGCGTGTTTACAGTGGTGTAGGTGGGAGCA 67445

Qy       6481 TATCTCGCCAATGGGAATGGCTGATGAATCCACACTTAGTGCTCGGATCAGGTGCTCTGT 6540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67446 TATCTCGCCAATGGGAATGGCTGATGAATCCACACTTAGTGCTCGGATCAGGTGCTCTGT 67505

Qy       6541 GTACTTTACTAGTGCCGATGAACCTAAATTTGGCGTGTTTAGGTCGTACCAACCAAAAAA 6600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67506 GTACTTTACTAGTGCCGATGAACCTAAATTTGGCGTGTTTAGGTCGTACCAACCAAAAAA 67565

Qy       6601 CTGCCTGGCCTCGCCTTTTAGAAGCTCTATGCAAGCTTTAATCGTGTCCAATAGCTTGTC 6660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67566 CTGCCTGGCCTCGCCTTTTAGAAGCTCTATGCAAGCTTTAATCGTGTCCAATAGCTTGTC 67625

Qy       6661 TTTAAAAATGCAACTCTGGTACATCTTTATGACAGTGGTCCAAAAAAAACAAAGATTTAA 6720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67626 TTTAAAAATGCAACTCTGGTACATCTTTATGACAGTGGTCCAAAAAAAACAAAGATTTAA 67685

Qy       6721 AAACACGTTAAACTGCGTGTCCTGGTAGTCCTCGTATATAACCTCCTCAACAAGAAAAAA 6780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67686 AAACACGTTAAACTGCGTGTCCTGGTAGTCCTCGTATATAACCTCCTCAACAAGAAAAAA 67745

Qy       6781 TTTTTTAACCAAACTCGCCAGGTACTGAAACGATGGTACGGACAGGTCGTGAAAAGTGTC 6840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67746 TTTTTTAACCAAACTCGCCAGGTACTGAAACGATGGTACGGACAGGTCGTGAAAAGTGTC 67805

Qy       6841 TATCATCTCCTTCACCTCCCCTTTGAAGGTTTTGGTTACACCGACCATGTGAGATAGGCA 6900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67806 TATCATCTCCTTCACCTCCCCTTTGAAGGTTTTGGTTACACCGACCATGTGAGATAGGCA 67865

Qy       6901 CCAGCTCAGGGGGGAGGTCGGATCGTGAGGGGGTAACAATTCGTGGGCGTGGGGGTACTC 6960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67866 CCAGCTCAGGGGGGAGGTCGGATCGTGAGGGGGTAACAATTCGTGGGCGTGGGGGTACTC 67925

Qy       6961 CAGTTTTAGGATCTCCGGCAGGACGCGTACGAGTTCTTCGTCAAATCCTACGCAGCCCGC 7020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67926 CAGTTTTAGGATCTCCGGCAGGACGCGTACGAGTTCTTCGTCAAATCCTACGCAGCCCGC 67985

Qy       7021 ATCCACCAGGGGCAGTACTGAGTTAATAAGTCCCCGAATTCTATCGTTCGTAATTTTGTG 7080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67986 ATCCACCAGGGGCAGTACTGAGTTAATAAGTCCCCGAATTCTATCGTTCGTAATTTTGTG 68045

Qy       7081 GAGCTCCTTTAGACAGTAAACCATGTTGCTCCGACACTGCGGTACGATGAACCGCTCTCC 7140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68046 GAGCTCCTTTAGACAGTAAACCATGTTGCTCCGACACTGCGGTACGATGAACCGCTCTCC 68105

Qy       7141 TTGTCTGTGTTGCATGAGCACCTGGAGTATCTTAGACAGGTATAGGGCCCGTTCCACTCT 7200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68106 TTGTCTGTGTTGCATGAGCACCTGGAGTATCTTAGACAGGTATAGGGCCCGTTCCACTCT 68165

Qy       7201 AAGGGTTGCCGCTTGGTAGAGCGGATTGCGGGCATTGACGGTCAGTGTGTTCATTTCACC 7260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68166 AAGGGTTGCCGCTTGGTAGAGCGGATTGCGGGCATTGACGGTCAGTGTGTTCATTTCACC 68225

Qy       7261 GACTATCATGCCGAGTGGCGGATAGTGGAACTTGTACATATGATTTAAAAGATGGTCTTT 7320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68226 GACTATCATGCCGAGTGGCGGATAGTGGAACTTGTACATATGATTTAAAAGATGGTCTTT 68285

Qy       7321 CCCGTAATGGCGACTCATGGTCGTGTTTTAACGTTCACAGGGTGTTAAAAAACTCAGAAG 7380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68286 CCCGTAATGGCGACTCATGGTCGTGTTTTAACGTTCACAGGGTGTTAAAAAACTCAGAAG 68345

Qy       7381 GTCCTGCGTGGACAACTTTTGCGGACTCTCCGAACCGCTCAGGGATAAGCTGTACCACGC 7440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68346 GTCCTGCGTGGACAACTTTTGCGGACTCTCCGAACCGCTCAGGGATAAGCTGTACCACGC 68405
```

```
Qy       7441 GGTGCTTCTTATAAACGAAGCATACGTCAAAAAACATGACCCACGTGACCTAAATAAGCT 7500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68406 GGTGCTTCTTATAAACGAAGCATACGTCAAAAAACATGACCCACGTGACCTAAATAAGCT 68465

Qy       7501 CACAGAAGAAGTCTGTCGATGTATTGTAATGGAATGTGCCTCCCTGGGTCCCATCTCCGG 7560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68466 CACAGAAGAAGTCTGTCGATGTATTGTAATGGAATGTGCCTCCCTGGGTCCCATCTCCGG 68525

Qy       7561 GCTCATCGCGGATCTGAATCTCTTTAACCTGTTCTGCCTCTACCGGGGTTCCCGGGTTAA 7620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68526 GCTCATCGCGGATCTGAATCTCTTTAACCTGTTCTGCCTCTACCGGGGTTCCCGGGTTAA 68585

Qy       7621 AACCCGCGGGGCCGCCACCTGTAACGTCCCGTGCGCAGAATGCGCCCAGGGCATCGTGAG 7680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68586 AACCCGCGGGGCCGCCACCTGTAACGTCCCGTGCGCAGAATGCGCCCAGGGCATCGTGAG 68645

Qy       7681 AATTCTCACCGAGAGGGCCCTTTGTTGTACCGAAAAGATGTTCATAGCGTCTGCCTGCAG 7740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68646 AATTCTCACCGAGAGGGCCCTTTGTTGTACCGAAAAGATGTTCATAGCGTCTGCCTGCAG 68705

Qy       7741 CGGGGTCGTTATACCGCCCCAGTTGGCCAGAGTTCTGCACGACGTCTACGCGGAAATGAA 7800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68706 CGGGGTCGTTATACCGCCCCAGTTGGCCAGAGTTCTGCACGACGTCTACGCGGAAATGAA 68765

Qy       7801 GGCCAAGTGTTTGGGGGCGTGGCGTCGTCTCATATGTTGCAGGCGGCCTATCATGGCTAT 7860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68766 GGCCAAGTGTTTGGGGGCGTGGCGTCGTCTCATATGTTGCAGGCGGCCTATCATGGCTAT 68825

Qy       7861 TGCCGACTCGGTCCTCGTCACTTATAATACCCTGGATGCCGAGGGAAAACTGGAACTTAG 7920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68826 TGCCGACTCGGTCCTCGTCACTTATAATACCCTGGATGCCGAGGGAAAACTGGAACTTAG 68885

Qy       7921 ACTCAAAGCACTGTGCAAACTTGTTTTTCAACCCATCTTTCTTCAAAGAATCTTAGCCCC 7980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68886 ACTCAAAGCACTGTGCAAACTTGTTTTTCAACCCATCTTTCTTCAAAGAATCTTAGCCCC 68945

Qy       7981 TATGCAGCTACTGGCCAACGGGAAGATGGTTCCTGACAATTATTTTACCATCACCGGTAC 8040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68946 TATGCAGCTACTGGCCAACGGGAAGATGGTTCCTGACAATTATTTTACCATCACCGGTAC 69005

Qy       8041 GGCCGAGAAGAGGCGCCCTGTCGTGACTGGAAGTACTAGCGGGATGACGTGTCCGGGAAG 8100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69006 GGCCGAGAAGAGGCGCCCTGTCGTGACTGGAAGTACTAGCGGGATGACGTGTCCGGGAAG 69065

Qy       8101 CAGCCTTGTCCCCGATTCCTTAATCCTGCCAGTATGCGAACCGGGGCTTCTCCCGGCACC 8160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69066 CAGCCTTGTCCCCGATTCCTTAATCCTGCCAGTATGCGAACCGGGGCTTCTCCCGGCACC 69125

Qy       8161 CCTGGTTGACCTCAGTAATGTCTTAGAAAATCCAGAAATCATCCTCAGCGCCCCACCCCT 8220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69126 CCTGGTTGACCTCAGTAATGTCTTAGAAAATCCAGAAATCATCCTCAGCGCCCCACCCCT 69185

Qy       8221 GAGTCAATTTGTCATCACAAACACGCACCCCAGTCTGCCTCAGTCAGTCAGCATTATTAC 8280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69186 GAGTCAATTTGTCATCACAAACACGCACCCCAGTCTGCCTCAGTCAGTCAGCATTATTAC 69245

Qy       8281 GCCAACCCAGGGCGTTGTTCCCGGCCAATGTTTTATGGACACGTGGAAAGCGGTGTCACA 8340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69246 GCCAACCCAGGGCGTTGTTCCCGGCCAATGTTTTATGGACACGTGGAAAGCGGTGTCACA 69305

Qy       8341 GAGCATTCACCACCAGGCACAGACGCCTATTTTGGCCGCCGCACTAACCGGTTCGACATC 8400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69306 GAGCATTCACCACCAGGCACAGACGCCTATTTTGGCCGCCGCACTAACCGGTTCGACATC 69365

Qy       8401 TGCGGCCCCTGGCCCGCATATCGCATGTTCCCCAGTTGCCGGCACGTCTCGGCAGGTGGA 8460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69366 TGCGGCCCCTGGCCCGCATATCGCATGTTCCCCAGTTGCCGGCACGTCTCGGCAGGTGGA 69425

Qy       8461 AGGGTCCGCGGGCGTCGATTGCGGGAAACCAGCATGCGTTCCGCAGCCCGCGTTACCGCC 8520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69426 AGGGTCCGCGGGCGTCGATTGCGGGAAACCAGCATGCGTTCCGCAGCCCGCGTTACCGCC 69485

Qy       8521 CAATGTCCCCGCCAAGAGGATGGAAACGGTAGCACAGTTGGGAAACGCTCCCGTAAAAAA 8580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69486 CAATGTCCCCGCCAAGAGGATGGAAACGGTAGCACAGTTGGGAAACGCTCCCGTAAAAAA 69545

Qy       8581 CGTCCACATCGGAGGCCGCGTATACGCTCCACTGGTTAATATACCAATAATAGACTTAAC 8640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69546 CGTCCACATCGGAGGCCGCGTATACGCTCCACTGGTTAATATACCAATAATAGACTTAAC 69605

Qy       8641 GTCCCCGTCAGGGTCCGGCCAGAGTCCGGCCGATATCGCCAACACTCCAGAGTCCCGCAT 8700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69606 GTCCCCGTCAGGGTCCGGCCAGAGTCCGGCCGATATCGCCAACACTCCAGAGTCCCGCAT 69665

Qy       8701 GGCGGCCGGCTCTCCGCCCTTCGCCGAAACCGCCGCAACGGTCCCCGCTAAGAGAAAGCA 8760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69666 GGCGGCCGGCTCTCCGCCCTTCGCCGAAACCGCCGCAACGGTCCCCGCTAAGAGAAAGCA 69725

Qy       8761 GCCACGCGAGGACGTGGCAGACAAAAGACTGAAGGGCGACGTTCGGGGCGCCGCAACAGT 8820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69726 GCCACGCGAGGACGTGGCAGACAAAAGACTGAAGGGCGACGTTCGGGGCGCCGCAACAGT 69785

Qy       8821 AAACCACCCTTTCCCGGGACCGTCCGGGATGCGCGTTCGCGAGCAGGGCTTATTCGATTT 8880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69786 AAACCACCCTTTCCCGGGACCGTCCGGGATGCGCGTTCGCGAGCAGGGCTTATTCGATTT 69845

Qy       8881 AATCGAAAGCTCCACGGATGTAACCGCGAACGCATCTGGACCCAAAAACGACGACGACAT 8940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69846 AATCGAAAGCTCCACGGATGTAACCGCGAACGCATCTGGACCCAAAAACGACGACGACAT 69905

Qy       8941 GCTAGCGGCTATCCTACAGGACCTGTATGGACTACAGTCCCCCCCGGCCATCGATTCCCC 9000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69906 GCTAGCGGCTATCCTACAGGACCTGTATGGACTACAGTCCCCCCCGGCCATCGATTCCCC 69965

Qy       9001 CTCCAGCAACTCGGACAATGAGGAGATATTTCCAGAGGTTAGTCCGCCATCTAGCGGCCA 9060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69966 CTCCAGCAACTCGGACAATGAGGAGATATTTCCAGAGGTTAGTCCGCCATCTAGCGGCCA 70025

Qy       9061 CGGATCGCCTTGA 9073
             |||||||||||||
Db      70026 CGGATCGCCTTGA 70038
```

RESULT 3
ABL33920/c
ID ABL33920 standard; DNA; 12590 BP.
XX
AC ABL33920;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1893.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;

```
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 1893; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 12590 BP; 2516 A; 310 C; 3233 G; 6531 T; 0 other;

  Query Match             0.2%;  Score 22;  DB 24;  Length 12590;
  Best Local Similarity   100.0%;  Pred. No. 8.9;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2363 TCATCTACTAAATAAATAAAAA 2384
             ||||||||||||||||||||||
Db       5900 TCATCTACTAAATAAATAAAAA 5879


RESULT 4
ABZ09971/c
ID   ABZ09971 standard; DNA; 3628 BP.
XX
AC   ABZ09971;
XX
DT   16-JAN-2003  (first entry)
XX
DE   Haematopoietic cell proliferation disorder related DNA sequence #111.
XX
KW   Human; haematopoietic cell proliferation disorder; cytostatic;
KW   gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW   cytosine methylation state; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200277272-A2.
XX
PD   03-OCT-2002.
XX
PF   26-MAR-2002; 2002WO-EP03401.
XX
PR   26-MAR-2001; 2001US-278333P.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Berlin K,  Braun A,  Distler J,  Guetig D,  Howe A,  Mueller J;
PI   Olek A,  Piepenbrock C,  Adorjan P,  Grabs G,  Lesche R,  Leu E;
PI   Lewin A,  Lipscher E,  Maier S,  Model F,  Mueller V,  Otto T;
PI   Pelet C,  Schwope I,  Ziebarth H;
XX
DR   WPI; 2003-018942/01.
XX
PT   Detecting and differentiating between hematopoietic cell proliferative
PT   disorders, comprises contacting a target nucleic acid with a reagent
PT   that distinguishes between methylated and non-methylated CpG
PT   dinucleotides -
XX
PS   Claim 28; SEQ ID 111; 117pp; English.
XX
CC   The present invention describes a method for detecting and
CC   differentiating between haematopoietic cell proliferative disorders
CC   associated with at least 1 gene and/or their regulatory regions in a
CC   subject. The method comprises contacting a target nucleic acid in a
CC   biological sample obtained from the subject with at least 1 reagent,
CC   which distinguishes between methylated and non-methylated CpG
CC   dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC   represent specifically claimed nucleotide sequences from the present
CC   invention. Oligonucleotides from the present invention can be used: for
CC   differentiating between healthy haematopoietic cells and proliferative
CC   disorder haematopoietic cells; for differentiating between acute
CC   lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC   determining the cytosine methylation state and/or single nucleotide
CC   polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC   related sequences and their complements; and as primers for the
CC   amplification of haematopoietic cell proliferation disorder related
CC   DNA sequences. The nucleotide sequences from the present invention can
CC   also be used for detecting a predisposition to, differentiation between
CC   subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC   haematopoietic cell proliferative disorders. The present method enables
CC   a highly specific classification of haematopoietic cell proliferative
CC   disorders allowing for improved and informed treatment of patients.
XX
SQ   Sequence 3628 BP; 734 A; 246 C; 1028 G; 1620 T; 0 other;

  Query Match             0.2%;  Score 21;  DB 25;  Length 3628;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2367 CTACTAAATAAATAAAAACAA 2387
             |||||||||||||||||||||
Db       1013 CTACTAAATAAATAAAAACAA 993


RESULT 5
ABZ10117/c
ID   ABZ10117 standard; DNA; 3628 BP.
XX
AC   ABZ10117;
XX
DT   16-JAN-2003  (first entry)
XX
DE   Haematopoietic cell proliferation disorder related DNA sequence #257.
XX
KW   Human; haematopoietic cell proliferation disorder; cytostatic;
KW   gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW   cytosine methylation state; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200277272-A2.
XX
PD   03-OCT-2002.
XX
PF   26-MAR-2002; 2002WO-EP03401.
XX
PR   26-MAR-2001; 2001US-278333P.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Berlin K,  Braun A,  Distler J,  Guetig D,  Howe A,  Mueller J;
PI   Olek A,  Piepenbrock C,  Adorjan P,  Grabs G,  Lesche R,  Leu E;
PI   Lewin A,  Lipscher E,  Maier S,  Model F,  Mueller V,  Otto T;
PI   Pelet C,  Schwope I,  Ziebarth H;
XX
DR   WPI; 2003-018942/01.
XX
PT   Detecting and differentiating between hematopoietic cell proliferative
PT   disorders, comprises contacting a target nucleic acid with a reagent
PT   that distinguishes between methylated and non-methylated CpG
PT   dinucleotides -
XX
PS   Claim 28; SEQ ID 257; 117pp; English.
XX
CC   The present invention describes a method for detecting and
CC   differentiating between haematopoietic cell proliferative disorders
CC   associated with at least 1 gene and/or their regulatory regions in a
```

```
CC  subject. The method comprises contacting a target nucleic acid in a
CC  biological sample obtained from the subject with at least 1 reagent,
CC  which distinguishes between methylated and non-methylated CpG
CC  dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC  represent specifically claimed nucleotide sequences from the present
CC  invention. Oligonucleotides from the present invention can be used: for
CC  differentiating between healthy haematopoietic cells and proliferative
CC  disorder haematopoietic cells; for differentiating between acute
CC  lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC  determining the cytosine methylation state and/or single nucleotide
CC  polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC  related sequences and their complements; and as primers for the
CC  amplification of haematopoietic cell proliferation disorder related
CC  DNA sequences. The nucleotide sequences from the present invention can
CC  also be used for detecting a predisposition to, differentiation between
CC  subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC  haematopoietic cell proliferative disorders. The present method enables
CC  a highly specific classification of haematopoietic cell proliferative
CC  disorders allowing for improved and informed treatment of patients.
XX
SQ  Sequence 3628 BP; 734 A; 0 C; 1028 G; 1866 T; 0 other;

  Query Match             0.2%;  Score 21;  DB 25;  Length 3628;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2367 CTACTAAATAAATAAAAACAA 2387
             |||||||||||||||||||||
Db       1013 CTACTAAATAAATAAAAACAA 993


RESULT 6
ABL32593/c
ID  ABL32593 standard; DNA; 6127 BP.
XX
AC  ABL32593;
XX
DT  26-MAR-2002  (first entry)
XX
DE  Human immune system associated gene SEQ ID NO: 566.
XX
KW  Human; immune system disease; cytosine methylation; antiasthmatic;
KW  antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW  neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW  antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW  antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW  acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW  neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW  gene; ds.
XX
OS  Homo sapiens.
XX
PN  WO200200928-A2.
XX
PD  03-JAN-2002.
XX
PF  02-JUL-2001; 2001WO-EP07537.
XX
PR  30-JUN-2000; 2000DE-1032529.
PR  01-SEP-2000; 2000DE-1043826.
XX
PA  (EPIG-) EPIGENOMICS AG.
XX
PI  Olek A,  Piepenbrock C,  Berlin K;
XX
DR  WPI; 2002-130909/17.
XX
PT  Nucleic acid comprising fragment of chemically modified gene, useful
PT  for diagnosis and treatment of diseases associated with abnormal
PT  cytosine methylation  -
XX
PS  Claim 1; SEQ ID NO 566; 32pp + Sequence Listing; German.
XX
CC  The present invention provides a number of human immune system associated
CC  genes which are modified by the methylation of cytosines. The sequences
CC  can be used in the diagnosis and treatment of immune system disorders,
CC  including eye diseases such as retinopathy, neovascular glaucoma and
CC  macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC  leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC  rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC  diseases. The present sequence is a gene of the invention.
XX
SQ  Sequence 6127 BP; 1578 A; 67 C; 1387 G; 3095 T; 0 other;

  Query Match             0.2%;  Score 21;  DB 24;  Length 6127;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2367 CTACTAAATAAATAAAAACAA 2387
             |||||||||||||||||||||
Db       2821 CTACTAAATAAATAAAAACAA 2801


RESULT 7
ABL05364/c
ID  ABL05364 standard; cDNA; 36383 BP.
XX
AC  ABL05364;
XX
DT  26-MAR-2002  (first entry)
XX
DE  Drosophila melanogaster expressed polynucleotide SEQ ID NO 10574.
XX
KW  Drosophila; developmental biology; cell signalling; insecticide;
KW  pharmaceutical; gene; ss.
XX
OS  Drosophila melanogaster.
XX
PN  WO200171042-A2.
XX
PD  27-SEP-2001.
XX
PF  23-MAR-2001; 2001WO-US09231.
XX
PR  23-MAR-2000; 2000US-191637P.
PR  11-JUL-2000; 2000US-0614150.
XX
PA  (PEKE ) PE CORP NY.
XX
PI  Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR  WPI; 2001-656860/75.
DR  P-PSDB; ABB61261.
XX
PT  New isolated nucleic acid detection reagent for detecting 1000 or more
PT  genes from Drosophila and for elucidating cell signalling and cell-cell
PT  interactions -
XX
PS  Claim 1; SEQ ID NO 10574; 21pp + Sequence Listing; English.
XX
CC  The invention relates to an isolated nucleic acid detection reagent
CC  capable of detecting 1000 or more genes from Drosophila. The invention is
CC  useful in developmental biology and in elucidating cell signalling and
CC  cell-cell interactions in higher eukaryotes for the development of
CC  insecticides, therapeutics and pharmaceutical drugs. The invention
CC  discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC  sequences (ABL01840-ABL16175) and the encoded proteins
CC  (ABB57737-ABB72072).
CC  The sequence data for this patent did not form part of the printed
CC  specification, but was obtained in electronic format directly from WIPO
CC  at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ  Sequence 36383 BP; 10996 A; 6940 C; 6897 G; 11550 T; 0 other;

  Query Match             0.2%;  Score 21;  DB 23;  Length 36383;
  Best Local Similarity   100.0%;  Pred. No. 27;
```

```
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2371 TAAATAAATAAAAACAAGCAA 2391
              ||||||||||||||||||||
Db      11364 TAAATAAATAAAAACAAGCAA 11344


RESULT 8
AAH74201
ID   AAH74201 standard; DNA; 124884 BP.
XX
AC   AAH74201;
XX
DT   15-OCT-2001  (first entry)
XX
DE   Nucleotide sequence of the Varicella virus Dumas strain.
XX
KW   Dumas strain; pox vaccine; vaccine; poxvirus disease; chickenpox; ss.
XX
OS   Varicella virus.
XX
PN   WO200156600-A1.
XX
PD   09-AUG-2001.
XX
PF   31-JAN-2001; 2001WO-JP00678.
XX
PR   31-JAN-2000; 2000JP-0062734.
XX
PA   (OSAU ) UNIV OSAKA.
XX
PI   Gomi Y,  Sunamachi H,  Takahashi M,  Yamanishi K;
XX
DR   WPI; 2001-488845/53.
XX
PT   Method for checking absence of mutation at specific positions of
PT   varicella virus genome for quality control of attenuated live varicella
PT   vaccine  -
XX
PS   Claim 1; Page 61-105; 158pp; Japanese.
XX
CC   The present sequence represents the nucleotide sequence of Varicella
CC   virus, Dumas strain. The specification describes a method for the
CC   quality control of an attenuated pox vaccine, characterised in that
CC   the DNA sequence of specific parts of the viral genome in a sample is
CC   determined and proved to be conserved rather than mutated in comparison
CC   with the same parts of a reference viral genome. The method is useful
CC   for quality control of vaccines for use in the prevention of poxvirus
CC   diseases such as chickenpox.
XX
SQ   Sequence 124884 BP; 33789 A; 29295 C; 28177 G; 33623 T; 0 other;

  Query Match             0.2%;  Score 21;  DB 22;  Length 124884;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        711 CATAGTATGCGTCGGTTCGCC 731
              |||||||||||||||||||||
Db      96832 CATAGTATGCGTCGGTTCGCC 96852


RESULT 9
AAH74202
ID   AAH74202 standard; DNA; 125157 BP.
XX
AC   AAH74202;
XX
DT   15-OCT-2001  (first entry)
XX
DE   Nucleotide sequence of the Varicella virus Oka strain.
XX
KW   Oka strain; pox vaccine; vaccine; poxvirus disease; chickenpox; ss.
XX
OS   Varicella virus.
XX
PN   WO200156600-A1.
XX
PD   09-AUG-2001.
XX
PF   31-JAN-2001; 2001WO-JP00678.
XX
PR   31-JAN-2000; 2000JP-0062734.
XX
PA   (OSAU ) UNIV OSAKA.
XX
PI   Gomi Y,  Sunamachi H,  Takahashi M,  Yamanishi K;
XX
DR   WPI; 2001-488845/53.
XX
PT   Method for checking absence of mutation at specific positions of
PT   varicella virus genome for quality control of attenuated live varicella
PT   vaccine  -
XX
PS   Disclosure; Page 106-150; 158pp; Japanese.
XX
CC   The present sequence represents the nucleotide sequence of Varicella
CC   virus, Oka strain. The specification describes a method for the
CC   quality control of an attenuated pox vaccine, characterised in that
CC   the DNA sequence of specific parts of the viral genome in a sample is
CC   determined and proved to be conserved rather than mutated in comparison
CC   with the same parts of a reference viral genome. The method is useful
CC   for quality control of vaccines for use in the prevention of poxvirus
CC   diseases such as chickenpox.
XX
SQ   Sequence 125157 BP; 33776 A; 29439 C; 28285 G; 33624 T; 33 other;

  Query Match             0.2%;  Score 21;  DB 22;  Length 125157;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        711 CATAGTATGCGTCGGTTCGCC 731
              |||||||||||||||||||||
Db      96778 CATAGTATGCGTCGGTTCGCC 96798


RESULT 10
ABK80864/c
ID   ABK80864 standard; DNA; 255 BP.
XX
AC   ABK80864;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Bacillus clausii genomic sequence tag (GST) #3707.
XX
KW   Differential gene expression; genomic sequenced tag; GST;
KW   altered culture condition; environmental stress;
KW   physiological provocation; ds.
XX
OS   Bacillus clausii.
XX
PN   WO200229113-A2.
XX
PD   11-APR-2002.
XX
PF   05-OCT-2001; 2001WO-US31437.
XX
PR   06-OCT-2000; 2000US-0680598.
PR   27-MAR-2001; 2001US-279526P.
XX
PA   (NOVO ) NOVOZYMES BIOTECH INC.
PA   (NOVO ) NOVOZYMES AS.
XX
PI   Berka R,  Clausen IG;
XX
DR   WPI; 2002-416684/44.
```

```
XX
PT   Monitoring differential expression of several genes in first Bacillus
PT   cell relative to expression of same genes in one or more second
PT   Bacillus cells, by using substrate containing Bacillus genomic
PT   sequenced tag array  -
XX
PS   Claim 11; SEQ ID NO 8155; 200pp; English.
XX
CC   The invention describes a method of monitoring differential expression of
CC   genes in a first Bacillus cell relative to expression of the genes in
CC   other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC   isolated from Bacillus cells to a substrate containing array of Bacillus
CC   genomic sequenced tags (GST), examining the array, and determining
CC   relative gene expression by an observed hybridisation reporter signal of
CC   a spot in the array. The method is useful for measuring the expression of
CC   genes in a first Bacillus cell relative to expression of the same genes
CC   in one or more second Bacillus cells. The method is useful for monitoring
CC   global expression of several genes from a Bacillus cell, discovering new
CC   genes, identifying possible functions of unknown open reading frames and
CC   monitoring gene copy number variation and stability. Monitoring changes
CC   in expression of genes may be used to provide a representation of the way
CC   in which Bacillus cells adapt to changes in culture conditions,
CC   environmental stress or other physiological provocation. Extensive
CC   follow-up characterisation is unnecessary, when one spot on an array
CC   equals one gene or one open reading frame, since sequence information is
CC   available. This sequence represents a genomic sequence tag (GST) used in
CC   the method of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 255 BP; 71 A; 43 C; 68 G; 73 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 24;  Length 255;
  Best Local Similarity   100.0%;  Pred. No. 95;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6768 CAACAAGAAAAAATTTTTTA 6787
             ||||||||||||||||||||
Db         42 CAACAAGAAAAAATTTTTTA 23


RESULT 11
ABV62066/c
ID   ABV62066 standard; cDNA; 310 BP.
XX
AC   ABV62066;
XX
DT   13-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 62057.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 2000US-183319P.
PR   16-MAR-2000; 2000US-189862P.
PR   25-MAY-2000; 2000US-207454P.
PR   09-JUN-2000; 2000US-211314P.
PR   18-JUL-2000; 2000US-219007P.
PR   13-DEC-2000; 2000US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer  -
XX
PS   Claim 1; Page 11728; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for:
CC   (a) assessing whether a patient is afflicted with prostate cancer;
CC   (b) monitoring the progression of prostate cancer in a patient;
CC   (c) assessing the efficacy of a test compound to inhibit prostate
CC   cancer in a patient;
CC   (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC   in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound;
CC   (g) determining whether prostate cancer has metastasized in a patient;
CC   (h) assessing the aggressiveness or indolence of prostate cancer in a
CC   patient;
CC   (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ   Sequence 310 BP; 94 A; 42 C; 59 G; 115 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 23;  Length 310;
  Best Local Similarity   100.0%;  Pred. No. 94;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db         37 GCAAACATCAAATAATATAA 18


RESULT 12
ABK79195
ID   ABK79195 standard; DNA; 356 BP.
XX
AC   ABK79195;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Bacillus clausii genomic sequence tag (GST) #2038.
XX
KW   Differential gene expression; genomic sequenced tag; GST;
KW   altered culture condition; environmental stress;
KW   physiological provocation; ds.
XX
OS   Bacillus clausii.
XX
PN   WO200229113-A2.
XX
PD   11-APR-2002.
XX
PF   05-OCT-2001; 2001WO-US31437.
XX
PR   06-OCT-2000; 2000US-0680598.
PR   27-MAR-2001; 2001US-279526P.
XX
PA   (NOVO ) NOVOZYMES BIOTECH INC.
PA   (NOVO ) NOVOZYMES AS.
XX
PI   Berka R,  Clausen IG;
XX
DR   WPI; 2002-416684/44.
XX
PT   Monitoring differential expression of several genes in first Bacillus
PT   cell relative to expression of same genes in one or more second
PT   Bacillus cells, by using substrate containing Bacillus genomic
PT   sequenced tag array  -
```

```
XX
PS   Claim 11; SEQ ID NO 6486; 200pp; English.
XX
CC   The invention describes a method of monitoring differential expression of
CC   genes in a first Bacillus cell relative to expression of the genes in
CC   other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC   isolated from Bacillus cells to a substrate containing array of Bacillus
CC   genomic sequenced tags (GST), examining the array, and determining
CC   relative gene expression by an observed hybridisation reporter signal of
CC   a spot in the array. The method is useful for measuring the expression of
CC   genes in a first Bacillus cell relative to expression of the same genes
CC   in one or more second Bacillus cells. The method is useful for monitoring
CC   global expression of several genes from a Bacillus cell, discovering new
CC   genes, identifying possible functions of unknown open reading frames and
CC   monitoring gene copy number variation and stability. Monitoring changes
CC   in expression of genes may be used to provide a representation of the way
CC   in which Bacillus cells adapt to changes in culture conditions,
CC   environmental stress or other physiological provocation. Extensive
CC   follow-up characterisation is unnecessary, when one spot on an array
CC   equals one gene or one open reading frame, since sequence information is
CC   available. This sequence represents a genomic sequence tag (GST) used in
CC   the method of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 356 BP; 107 A; 62 C; 89 G; 97 T; 1 other;

  Query Match             0.2%;  Score 20;  DB 24;  Length 356;
  Best Local Similarity   100.0%;  Pred. No. 94;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        225 TGGTGCTGGGAAAAGTACCA 244
              ||||||||||||||||||||
Db         68 TGGTGCTGGGAAAAGTACCA 87


RESULT 13
ABA14638/c
ID   ABA14638 standard; DNA; 456 BP.
XX
AC   ABA14638;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 6969.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
```

```
PR  20-OCT-2000; 2000US-0241787.
PR  20-OCT-2000; 2000US-0241808.
PR  20-OCT-2000; 2000US-0241809.
PR  20-OCT-2000; 2000US-0241826.
PR  20-OCT-2000; 2000US-0242221.
PR  01-NOV-2000; 2000US-0244617.
PR  08-NOV-2000; 2000US-0246474.
PR  08-NOV-2000; 2000US-0246475.
PR  08-NOV-2000; 2000US-0246476.
PR  08-NOV-2000; 2000US-0246477.
PR  08-NOV-2000; 2000US-0246478.
PR  08-NOV-2000; 2000US-0246523.
PR  08-NOV-2000; 2000US-0246524.
PR  08-NOV-2000; 2000US-0246525.
PR  08-NOV-2000; 2000US-0246526.
PR  08-NOV-2000; 2000US-0246527.
PR  08-NOV-2000; 2000US-0246528.
PR  08-NOV-2000; 2000US-0246532.
PR  08-NOV-2000; 2000US-0246609.
PR  08-NOV-2000; 2000US-0246610.
PR  08-NOV-2000; 2000US-0246611.
PR  08-NOV-2000; 2000US-0246613.
PR  17-NOV-2000; 2000US-0249207.
PR  17-NOV-2000; 2000US-0249208.
PR  17-NOV-2000; 2000US-0249209.
PR  17-NOV-2000; 2000US-0249210.
PR  17-NOV-2000; 2000US-0249211.
PR  17-NOV-2000; 2000US-0249212.
PR  17-NOV-2000; 2000US-0249213.
PR  17-NOV-2000; 2000US-0249214.
PR  17-NOV-2000; 2000US-0249215.
PR  17-NOV-2000; 2000US-0249216.
PR  17-NOV-2000; 2000US-0249217.
PR  17-NOV-2000; 2000US-0249218.
PR  17-NOV-2000; 2000US-0249244.
PR  17-NOV-2000; 2000US-0249245.
PR  17-NOV-2000; 2000US-0249264.
PR  17-NOV-2000; 2000US-0249265.
PR  17-NOV-2000; 2000US-0249297.
PR  17-NOV-2000; 2000US-0249299.
PR  17-NOV-2000; 2000US-0249300.
PR  01-DEC-2000; 2000US-0250391.
PR  01-DEC-2000; 2000US-0251160.
PR  05-DEC-2000; 2000US-0251030.
PR  05-DEC-2000; 2000US-0251988.
PR  05-DEC-2000; 2000US-0256719.
PR  06-DEC-2000; 2000US-0251479.
PR  08-DEC-2000; 2000US-0251856.
PR  08-DEC-2000; 2000US-0251868.
PR  08-DEC-2000; 2000US-0251869.
PR  08-DEC-2000; 2000US-0251989.
PR  08-DEC-2000; 2000US-0251990.
PR  11-DEC-2000; 2000US-0254097.
PR  05-JAN-2001; 2001US-0259678.
XX
PA  (HUMA-) HUMAN GENOME SCI INC.
XX
PI  Rosen CA,  Barash SC,  Ruben SM;
XX
DR  WPI; 2001-541565/60.
XX
PT  Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT  useful for preventing, diagnosing and/or treating nervous system
PT  cancers and metastases -
XX
PS  Disclosure; SEQ ID NO 6969; 1701pp + Sequence Listing; English.
XX
CC  The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC  (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC  medical conditions e.g. by protein or gene therapy. The genes are
CC  isolated from a range of human tissues disclosed in the specification.
CC  The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC  in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC  and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC  marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC  (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC  haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC  disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC  colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC  (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC  epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC  and parasitic infections.
CC  Note: The sequence data for this patent did not form part of the
CC  printed specification, but was obtained in electronic format directly
CC  from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ  Sequence 456 BP; 143 A; 63 C; 83 G; 167 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 22;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 93;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db         78 GCAAACATCAAATAATATAA 59


RESULT 14
ABN94325
ID  ABN94325 standard; DNA; 461 BP.
XX
AC  ABN94325;
XX
DT  13-AUG-2002  (first entry)
XX
DE  Gene #823 used to diagnose liver cancer.
XX
KW  Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW  metastatic liver tumour; cytostatic; expression profile; disease state;
KW  disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS  Homo sapiens.
XX
PN  WO200229103-A2.
XX
PD  11-APR-2002.
XX
PF  02-OCT-2001; 2001WO-US30589.
XX
PR  02-OCT-2000; 2000US-237054P.
XX
PA  (GENE-) GENE LOGIC INC.
XX
PI  Horne D,  Alvares C,  Peres-Da-Silva S,  Vockley JG;
XX
DR  WPI; 2002-426119/45.
XX
PT  Diagnosing and detecting the progression of liver cancer,
PT  hepatocellular carcinoma or metastatic liver tumor in a patient,
PT  involves detecting the level of expression of two or more genes in a
PT  liver tissue sample -
XX
PS  Claim 1; SEQ ID NO 823; 298pp; English.
XX
CC  The invention relates to a novel method for diagnosing and detecting the
CC  progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC  tumour in a patient, and differentiating metastatic liver cancer from
CC  hepatocellular carcinoma in a patient, involving detecting the level of
CC  expression of two or more genes represented in ABN93503-ABN97455 in a
CC  tissue sample. The method of the invention has hepatotropic, and
CC  cytostatic activity. The method is useful for diagnosing and detecting
CC  the progression of liver cancer, hepatocellular carcinoma and metastatic
CC  liver carcinoma in a patient. The method is useful for identifying
CC  expression profiles which serve as useful diagnostic markers as well as
CC  markers that can be used to monitor disease states, disease progression,
CC  drug toxicity, drug efficacy and drug metabolism.
```

```
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 461 BP; 162 A; 85 C; 62 G; 152 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 24;  Length 461;
  Best Local Similarity   100.0%;  Pred. No. 93;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
              ||||||||||||||||||||
Db        360 GCAAACATCAAATAATATAA 379


RESULT 15
ABA14637/c
ID   ABA14637 standard; DNA; 483 BP.
XX
AC   ABA14637;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 6968.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   20-OCT-2000; 2000US-0242221.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
```

```
PR  17-NOV-2000; 2000US-0249208.
PR  17-NOV-2000; 2000US-0249209.
PR  17-NOV-2000; 2000US-0249210.
PR  17-NOV-2000; 2000US-0249211.
PR  17-NOV-2000; 2000US-0249212.
PR  17-NOV-2000; 2000US-0249213.
PR  17-NOV-2000; 2000US-0249214.
PR  17-NOV-2000; 2000US-0249215.
PR  17-NOV-2000; 2000US-0249216.
PR  17-NOV-2000; 2000US-0249217.
PR  17-NOV-2000; 2000US-0249218.
PR  17-NOV-2000; 2000US-0249244.
PR  17-NOV-2000; 2000US-0249245.
PR  17-NOV-2000; 2000US-0249264.
PR  17-NOV-2000; 2000US-0249265.
PR  17-NOV-2000; 2000US-0249297.
PR  17-NOV-2000; 2000US-0249299.
PR  17-NOV-2000; 2000US-0249300.
PR  01-DEC-2000; 2000US-0250391.
PR  01-DEC-2000; 2000US-0251160.
PR  05-DEC-2000; 2000US-0251030.
PR  05-DEC-2000; 2000US-0251988.
PR  05-DEC-2000; 2000US-0256719.
PR  06-DEC-2000; 2000US-0251479.
PR  08-DEC-2000; 2000US-0251856.
PR  08-DEC-2000; 2000US-0251868.
PR  08-DEC-2000; 2000US-0251869.
PR  08-DEC-2000; 2000US-0251989.
PR  08-DEC-2000; 2000US-0251990.
PR  11-DEC-2000; 2000US-0254097.
PR  05-JAN-2001; 2001US-0259678.
XX
PA  (HUMA-) HUMAN GENOME SCI INC.
XX
PI  Rosen CA,  Barash SC,  Ruben SM;
XX
DR  WPI; 2001-541565/60.
XX
PT  Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT  useful for preventing, diagnosing and/or treating nervous system
PT  cancers and metastases -
XX
PS  Disclosure; SEQ ID NO 6968; 1701pp + Sequence Listing; English.
XX
CC  The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC  (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC  medical conditions e.g. by protein or gene therapy. The genes are
CC  isolated from a range of human tissues disclosed in the specification.
CC  The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC  in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC  and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC  marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC  (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC  haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC  disease, multiple sclerosis, rhéumatoid arthritis and ulcerative
CC  colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC  (d) wound healing; (e) neurological diseasés e.g. cerebral anoxia and
CC  epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC  and parasitic infections.
CC  Note: The sequence data for this patent did not form part of the
CC  printed specification, but was obtained in electronic format directly
CC  from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ  Sequence 483 BP; 155 A; 67 C; 86 G; 175 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 22;  Length 483;
  Best Local Similarity   100.0%;  Pred. No. 93;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
              ||||||||||||||||||||
Db        105 GCAAACATCAAATAATATAA 86
```

```
RESULT 16
AAI95250
ID  AAI95250 standard; cDNA; 736 BP.
XX
AC  AAI95250;
XX
DT  13-NOV-2001  (first entry)
XX
DE  Human neuroblastoma expressed polynucleotide SEQ ID NO 1325.
XX
KW  Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX
OS  Homo sapiens.
XX
PN  WO200166719-A1.
XX
PD  13-SEP-2001.
XX
PF  02-MAR-2001; 2001WO-JP01629.
XX
PR  07-MAR-2000; 2000JP-0159195.
XX
PA  (CHIB-) CHIBA PREFECTURE.
PA  (HISM ) HISAMITSU PHARM CO LTD.
XX
PI  Nakagawara A;
XX
DR  WPI; 2001-565584/63.
XX
PT  Nucleic acids originating in gene expressed in human neuroblastoma,
PT  useful as probe or primer in diagnosing prognosis of human
PT  neuroblastoma, malignancy and susceptibility indicator or tumour marker
PT  for anti-cancer agents -
XX
PS  Claim 1; Page 998; 2979pp; Japanese.
XX
CC  The invention relates to novel genes (AAI93926-AAI97963) expressed in
CC  human neuroblastoma. The nucleic acids are applicable as a probe or
CC  primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC  susceptibility indicators or tumour markers for anti-cancer agents. The
CC  gene information for diagnosing prognosis is related to factors similar
CC  to that for N-myc and TrkA genes.
XX
SQ  Sequence 736 BP; 262 A; 143 C; 97 G; 231 T; 3 other;

  Query Match             0.2%;  Score 20;  DB 22;  Length 736;
  Best Local Similarity   100.0%;  Pred. No. 92;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
              ||||||||||||||||||||
Db        333 GCAAACATCAAATAATATAA 352


RESULT 17
ABK71638
ID  ABK71638 standard; cDNA; 774 BP.
XX
AC  ABK71638;
XX
DT  30-JUL-2002  (first entry)
XX
DE  Human dithp polynucleotide #104.
XX
KW  Human; dithp; diagnostic and therapeutic polynucleotide; gene; ss; bone;
KW  cell proliferative disorder; cancer; tumour; autoimmune disorder; brain;
KW  inflammatory disorder; viral infection; bacterial infection; seizure;
KW  fungal infection; parasitic infections; developmental disorder; breast;
KW  endocrine disorder; metabolic disorder; neurological disorder; cervix;
KW  gastrointestinal disorder; transport disorder; gene therapy; kidney;
KW  adrenal gland; bone marrow; lung; ovary; pancreas; prostate; spleen;
KW  skin; testis; thymus.
```

```
XX
OS   Homo sapiens.
XX
PN   WO200220754-A2.
XX
PD   14-MAR-2002.
XX
PF   29-AUG-2001; 2001WO-US27127.
XX
PR   05-SEP-2000; 2000US-229747P.
PR   05-SEP-2000; 2000US-229748P.
PR   05-SEP-2000; 2000US-229749P.
PR   05-SEP-2000; 2000US-229750P.
PR   05-SEP-2000; 2000US-229751P.
PR   05-SEP-2000; 2000US-230583P.
PR   06-SEP-2000; 2000US-230505P.
PR   06-SEP-2000; 2000US-230514P.
PR   06-SEP-2000; 2000US-230515P.
PR   06-SEP-2000; 2000US-230517P.
PR   06-SEP-2000; 2000US-230518P.
PR   06-SEP-2000; 2000US-230519P.
PR   06-SEP-2000; 2000US-230595P.
PR   06-SEP-2000; 2000US-230597P.
PR   06-SEP-2000; 2000US-230598P.
PR   06-SEP-2000; 2000US-230599P.
PR   06-SEP-2000; 2000US-230610P.
PR   06-SEP-2000; 2000US-230865P.
PR   06-SEP-2000; 2000US-230988P.
PR   07-SEP-2000; 2000US-230951P.
PR   07-SEP-2000; 2000US-231163P.
PR   07-SEP-2000; 2000US-231167P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Stuart J,  Lincoln SE,  Altus CM,  Dufour GE,  Chalup MS,  Hillman JL;
PI   Jones AL,  Yu JY,  Wright RJ,  Gietzen D,  Liu TF,  Yap PE,  Dahl CR;
PI   Momiyama MG,  Bradley DL,  Rohatgi SD,  Harris B,  Roseberry AM;
PI   Gerstin EH,  Peralta CH,  David MH,  Panzer SR,  Flores V,  Daffo A;
PI   Marwaha R,  Chen AJ,  Chang SC,  Au AP,  Inman RR;
XX
DR   WPI; 2002-383054/41.
DR   P-PSDB; ABG60047.
XX
PT   An isolated polynucleotide useful in diagnostics and therapeutics -
XX
PS   Claim 1; Page 454; 686pp; English.
XX
CC   The invention relates to human diagnostic and therapeutic (dithp)
CC   polynucleotides and their associated polypeptides (DITHP polypeptides).
CC   The sequences of the invention are used in the treatment and diagnosis of
CC   cell proliferative disorders (e.g. atherosclerosis, cirrhosis), cancers
CC   (e.g. tumours of the adrenal gland, bone, bone marrow, brain, breast,
CC   cervix, kidney, lung, ovary, pancreas, prostate, skin, spleen, testis or
CC   thymus), autoimmune/inflammatory disorders (e.g. asthma, bronchitis,
CC   psoriasis, osteoporosis), viral infections, bacterial infections, fungal
CC   infections, parasitic infections, developmental disorders (e.g. anaemia,
CC   epilepsy), seizure disorders (e.g. cerebral palsy, spina bifida),
CC   endocrine disorders (e.g. thrombosis, aneurysm), metabolic disorders
CC   (e.g. obesity, diabetes), neurological disorders (e.g. stroke,
CC   amyotrophic lateral sclerosis, multiple sclerosis), gastrointestinal
CC   disorders (e.g. ulcerative colitis, lysinuria) and transport disorders
CC   (e.g. myotonic dystrophy, catatonia, peripheral neuropathy). Sequences
CC   ABK71535-ABK71809 represent human dithp polynucleotides of the invention.
XX
SQ   Sequence 774 BP; 280 A; 147 C; 136 G; 211 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 24;  Length 774;
  Best Local Similarity   100.0%;  Pred. No. 92;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4753 TACATACTTCTCATGTTTTA 4772
              ||||||||||||||||||||
Db        731 TACATACTTCTCATGTTTTA 750
```

```
RESULT 18
ABL91226
ID   ABL91226 standard; DNA; 834 BP.
XX
AC   ABL91226;
XX
DT   29-JUL-2002  (first entry)
XX
DE   Chlamydia pneumoniae cp7105 ORF DNA, SEQ ID NO:86.
XX
KW   Chlamydial infection; antigen; immunogen; vaccine; diagnosis;
KW   human respiratory disease; cardiovascular disease; atherosclerosis;
KW   coronary artery disease; carotid artery stenosis; myocardial infarction;
KW   cerebrovascular disease; aortic aneurysm; claudication; stroke;
KW   strain CWL029; open reading frame; ORF; gene; ds.
XX
OS   Chlamydia pneumoniae.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..834
FT                   /*tag=  a
FT                   /product= "cp7105"
XX
PN   WO200202606-A2.
XX
PD   10-JAN-2002.
XX
PF   03-JUL-2001; 2001WO-IB01445.
XX
PR   03-JUL-2000; 2000GB-0016363.
PR   11-JUL-2000; 2000GB-0017047.
PR   21-JUL-2000; 2000GB-0017983.
PR   07-AUG-2000; 2000GB-0019368.
PR   18-AUG-2000; 2000GB-0020440.
PR   14-SEP-2000; 2000GB-0022583.
PR   10-NOV-2000; 2000GB-0027549.
PR   22-DEC-2000; 2000GB-0031706.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Ratti G,  Grandi G;
XX
DR   WPI; 2002-154726/20.
DR   N-PSDB; ABB90568.
XX
PT   Novel Chlamydia pneumoniae protein useful in the manufacture of a
PT   medicament for treatment or prevention of infection due to Chlamydia,
PT   preferably Chlamydia pneumoniae, and for diagnostic purposes  -
XX
PS   Claim 5; Page 83; 364pp; English.
XX
CC   Sequences ABB90526-ABB90715 represent novel proteins from Chlamydia
CC   pneumoniae (strain CWL029), and ABL91184-ABL91373 represent DNA encoding
CC   them. The proteins are predicted to be immunogenic and may therefore be
CC   useful in vaccine production and for diagnostic purposes. Chlamydia
CC   pneumoniae is a common cause of respiratory disease in humans, and is
CC   also involved in the development of cardiovascular diseases such as
CC   atherosclerosis, coronary artery disease, carotid artery stenosis,
CC   myocardial infarction, cerebrovascular disease, aortic aneurysm,
CC   claudication and stroke. The proteins and nucleic acids of the invention
CC   may be used in vaccines and pharmaceutical compositions for the
CC   prevention or treatment of chlamydial infections, particularly Chlamydia
CC   pneumoniae infections. The proteins may also be used in the detection of
CC   Chlamydia pneumoniae, and the nucleic acids may be used in PCR, branched
CC   DNA probe assay or blotting techniques for determining Chlamydia
CC   pneumoniae gene expression. The present sequence represents a
CC   specifically claimed DNA which encodes a Chlamydia pneumoniae protein of
CC   the invention.
XX
SQ   Sequence 834 BP; 246 A; 158 C; 208 G; 222 T; 0 other;
```

```
  Query Match             0.2%;  Score 20;  DB 24;  Length 834;
  Best Local Similarity   100.0%;  Pred. No. 92;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6172 TTAACACCCTTAACTGGAAT 6191
              ||||||||||||||||||||
Db        391 TTAACACCCTTAACTGGAAT 410


RESULT 19
ABQ29020/c
ID   ABQ29020 standard; DNA; 848 BP.
XX
AC   ABQ29020;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 15611.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 848 BP; 121 A; 109 C; 285 G; 333 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 24;  Length 848;
  Best Local Similarity   100.0%;  Pred. No. 92;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8920 ACCCAAAAACGACGACGACA 8939
              ||||||||||||||||||||
Db        749 ACCCAAAAACGACGACGACA 730


RESULT 20
ABQ29021
ID   ABQ29021 standard; DNA; 848 BP.
XX
AC   ABQ29021;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 15612.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 848 BP; 333 A; 285 C; 109 G; 121 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 24;  Length 848;
  Best Local Similarity   100.0%;  Pred. No. 92;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8920 ACCCAAAAACGACGACGACA 8939
              ||||||||||||||||||||
```

```
Db        100 ACCCAAAAACGACGACGACA 119


RESULT 21
ABN74144
ID   ABN74144 standard; cDNA; 945 BP.
XX
AC   ABN74144;
XX
DT   03-JUL-2002  (first entry)
XX
DE   Bovine embryonic germ (EG) cell cDNA EST #195.
XX
KW   Bovine; Bos taurus; EST; expressed sequence tag; totipotence;
KW   development; gene; ss.
XX
OS   Bos taurus.
XX
PN   WO200194550-A2.
XX
PD   13-DEC-2001.
XX
PF   07-JUN-2001; 2001WO-US18576.
XX
PR   07-JUN-2000; 2000US-209874P.
PR   06-JUN-2001; 2001US-0876143.
XX
PA   (INFI-) INFIGEN INC.
XX
PI   Eilertsen KJ,  Pfister-Genskow M,  Childs L;
XX
DR   WPI; 2002-351289/38.
XX
PT   An expressed sequence tag (EST), the expression of which, or its
PT   complementary sequence, in a cell identifies the cell as a
PT   developmentally competent or incompetent cell -
XX
PS   Example 16; Page 334; 584pp; English.
XX
CC   The present invention describes an expressed sequence tag (EST), where
CC   the EST is an isolated, enriched, or purified nucleic acid sequence
CC   representing all or part of a gene, the expression of which, or its
CC   complementary sequence, in a cell identifies the cell as a
CC   developmentally competent or incompetent cell. Molecules which induce
CC   developmental competence in a cell line are useful for inducing
CC   totipotence in one or more cells. Molecules which induce developmental
CC   incompetence in a cell line are useful for preventing a full term
CC   pregnancy in an animal and inhibiting totipotence. The molecules are
CC   also useful for treating a disease in an animal by inducing development
CC   of one or more cells of the animal into a specific cell type. The
CC   present sequence represents a bovine EST which is given in the
CC   exemplification of the present invention.
XX
SQ   Sequence 945 BP; 193 A; 271 C; 235 G; 210 T; 36 other;

  Query Match              0.2%;  Score 20;  DB 24;  Length 945;
  Best Local Similarity   100.0%;  Pred. No. 91;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3217 CGCTGTTCTCCGAATGAGAG 3236
             ||||||||||||||||||||
Db        249 CGCTGTTCTCCGAATGAGAG 268


RESULT 22
AAZ45345/c
ID   AAZ45345 standard; DNA; 960 BP.
XX
AC   AAZ45345;
XX
DT   27-MAR-2000  (first entry)
XX
DE   DNA encoding a Bacillus sp. mannanase enzyme.
XX
KW   Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW   endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
KW   mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
KW   synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
KW   plant material degradation; recycled waste paper; paper making pulp;
KW   guar; locust bean gum; mannan-containing food; coffee extract;
KW   cleaning composition; machine washing; hard-surface cleaner;
KW   dishwashing; oral; dental; contact lens; body-care composition;
KW   fabric softener; oil well drilling; subterranean formation fracture; ss.
XX
OS   Bacillus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..960
FT                   /*tag=  a
FT                   /product= "mannanase"
FT                   /note= "not termination codon given"
FT   sig_peptide     1..84
FT                   /*tag=  b
XX
PN   WO9964619-A2.
XX
PD   16-DEC-1999.
XX
PF   10-JUN-1999;   99WO-DK00314.
XX
PR   10-JUN-1998;   98US-0111256.
PR   20-OCT-1998;   98DK-0001340.
PR   20-OCT-1998;   98DK-0001341.
PR   28-OCT-1998;   98US-0105970.
PR   28-OCT-1998;   98US-0106054.
PR   23-DEC-1998;   98DK-0001725.
PR   05-MAR-1999;   99DK-0000306.
PR   05-MAR-1999;   99DK-0000307.
PR   05-MAR-1999;   99DK-0000308.
PR   05-MAR-1999;   99DK-0000309.
PR   09-MAR-1999;   99US-0123543.
PR   10-MAR-1999;   99US-0123623.
PR   10-MAR-1999;   99US-0123641.
PR   11-MAR-1999;   99US-0123642.
XX
PA   (NOVO ) NOVO-NORDISK AS.
XX
PI   Kauppinen MS,  Schuelein M,  Schnorr K,  Andersen LN,  Bjornvad ME;
XX
DR   WPI; 2000-105891/09.
DR   P-PSDB; AAY54132.
XX
PT   New mannanases for treatment of textiles, plant material and coffee
PT   extract, and in cleaning compositions  -
XX
PS   Disclosure; Page 227-228; 242pp; English.
XX
CC   The present sequence encodes a mannanase enzyme (also known as
CC   mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
CC   The mannanase hydrolyses galactomannans. Specifically, mannanases
CC   hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
CC   glucomannans, and galactoglucomannans. The mannanase protein, or
CC   preparations containing it, are used to improve properties of cellulosic
CC   or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
CC   sizes or printing pastes). They are also used to degrade or modify
CC   plant materials (particularly recycled waste paper, paper making pulps,
CC   or material containing guar or locust bean gums (thickeners), or to
CC   reduce viscosity of mannan-containing foods or feeds). The mannanases
CC   are also used to process coffee extracts (to inhibit gel formation); in
CC   cleaning compositions (for machine washing of fabrics, as hard-surface
CC   cleaners, for hand or machine dishwashing, also in oral, dental, contact
CC   lens or body-care compositions) where they remove mannan-containing
CC   soils and prevent binding of some soils to cellulosics; and in fabric
CC   softeners. They can also be used in oil well drilling to fracture
CC   subterranean formations.
XX
```

```
SQ   Sequence 960 BP; 257 A; 183 C; 280 G; 240 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 21;  Length 960;
  Best Local Similarity   100.0%;  Pred. No. 91;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2392 ATGGAAAACATTGTGTTTTT 2411
              ||||||||||||||||||||
Db        653 ATGGAAAACATTGTGTTTTT 634


RESULT 23
ABK80177/c
ID   ABK80177 standard; DNA; 963 BP.
XX
AC   ABK80177;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Bacillus clausii genomic sequence tag (GST) #3020.
XX
KW   Differential gene expression; genomic sequenced tag; GST;
KW   altered culture condition; environmental stress;
KW   physiological provocation; ds.
XX
OS   Bacillus clausii.
XX
PN   WO200229113-A2.
XX
PD   11-APR-2002.
XX
PF   05-OCT-2001; 2001WO-US31437.
XX
PR   06-OCT-2000; 2000US-0680598.
PR   27-MAR-2001; 2001US-279526P.
XX
PA   (NOVO ) NOVOZYMES BIOTECH INC.
PA   (NOVO ) NOVOZYMES AS.
XX
PI   Berka R,  Clausen IG;
XX
DR   WPI; 2002-416684/44.
XX
PT   Monitoring differential expression of several genes in first Bacillus
PT   cell relative to expression of same genes in one or more second
PT   Bacillus cells, by using substrate containing Bacillus genomic
PT   sequenced tag array  -
XX
PS   Claim 11; SEQ ID NO 7468; 200pp; English.
XX
CC   The invention describes a method of monitoring differential expression of
CC   genes in a first Bacillus cell relative to expression of the genes in
CC   other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC   isolated from Bacillus cells to a substrate containing array of Bacillus
CC   genomic sequenced tags (GST), examining the array, and determining
CC   relative gene expression by an observed hybridisation reporter signal of
CC   a spot in the array. The method is useful for measuring the expression of
CC   genes in a first Bacillus cell relative to expression of the same genes
CC   in one or more second Bacillus cells. The method is useful for monitoring
CC   global expression of several genes from a Bacillus cell, discovering new
CC   genes, identifying possible functions of unknown open reading frames and
CC   monitoring gene copy number variation and stability. Monitoring changes
CC   in expression of genes may be used to provide a representation of the way
CC   in which Bacillus cells adapt to changes in culture conditions,
CC   environmental stress or other physiological provocation. Extensive
CC   follow-up characterisation is unnecessary, when one spot on an array
CC   equals one gene or one open reading frame, since sequence information is
CC   available. This sequence represents a genomic sequence tag (GST) used in
CC   the method of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 963 BP; 259 A; 181 C; 281 G; 242 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 24;  Length 963;
  Best Local Similarity   100.0%;  Pred. No. 91;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2392 ATGGAAAACATTGTGTTTTT 2411
              ||||||||||||||||||||
Db        653 ATGGAAAACATTGTGTTTTT 634


RESULT 24
ABQ91912/c
ID   ABQ91912 standard; cDNA; 1472 BP.
XX
AC   ABQ91912;
XX
DT   30-SEP-2002  (first entry)
XX
DE   Human NF-kB activating gene SEQ ID NO 2.
XX
KW   Human; NF-kB; nuclear factor kappa B; mouse; antiinflammatory;
KW   immunomodulator; cytostatic; antiinfective; osteopathic; nootropic;
KW   neuroprotective; anti-HIV; autoimmune disease; cancer; infection;
KW   bone disease; AIDS; neurodegenerative disease; ischaemic disorder;
KW   gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200253737-A1.
XX
PD   11-JUL-2002.
XX
PF   25-DEC-2001; 2001WO-JP11389.
XX
PR   28-DEC-2000; 2000JP-0402288.
PR   26-MAR-2001; 2001JP-0088912.
PR   24-AUG-2001; 2001JP-0254018.
XX
PA   (ASAH ) ASAHI KASEI KOGYO KK.
XX
PI   Matsuda A,  Honda G,  Muramatsu S,  Nagano Y;
XX
DR   WPI; 2002-583617/62.
DR   P-PSDB; ABP61424.
XX
PT   NF-approximatelykB activating gene and expressed protein, applicable in
PT   diagnosis and screening inhibitors or promoters to control excessive
PT   activation or inhibition for treating e.g. inflammations, autoimmune
PT   diseases and cancer  -
XX
PS   Claim 4; Page 101-104; 841pp; Japanese.
XX
CC   The invention relates to a purified protein (I), comprising one of 90
CC   fully defined sequences (ABP61424-ABP61513) or a protein based on any of
CC   the sequences but with some amino acids deleted, substituted or added and
CC   with a NF-kB (nuclear factor kappa B) activating effect. The protein and
CC   encoding gene (ABQ91912-ABQ92001) are useful in diagnosis and screening
CC   inhibitors or promoters to control excessive activation or inhibition
CC   and for treating e.g. inflammations, autoimmune diseases, cancers,
CC   infections, bone diseases, AIDS, neurodegenerative diseases or ischaemic
CC   disorders.
XX
SQ   Sequence 1472 BP; 421 A; 264 C; 324 G; 463 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 24;  Length 1472;
  Best Local Similarity   100.0%;  Pred. No. 90;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
              ||||||||||||||||||||
Db       1132 GCAAACATCAAATAATATAA 1113
```

```
RESULT 25
AAC61881/c
ID   AAC61881 standard; cDNA; 1839 BP.
XX
AC   AAC61881;
XX
DT   06-MAR-2001  (first entry)
XX
DE   cDNA encoding a human secreted protein.
XX
KW   Secreted protein; platelet disorder; stem cell disorder; osteoporosis;
KW   osteoarthritis; burn; incision; ulcer; periodontal disease; neuropathy;
KW   nervous system disease; bone growth; cosmetic plastic surgery;
KW   gut protection; gut regeneration; fibrosis; cancer;
KW   bone marrow transplantation; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             40..1059
FT                   /*tag=  a
FT                   /product= "secreted protein"
XX
PN   WO200061755-A2.
XX
PD   19-OCT-2000.
XX
PF   10-APR-2000; 2000WO-US09555.
XX
PR   09-APR-1999;   99US-0128574.
PR   20-AUG-1999;   99US-0150054.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Garcia PD;
XX
DR   WPI; 2000-665133/64.
DR   P-PSDB; AAB19395.
XX
PT   Novel secreted human proteins useful for stimulating blood cell
PT   generation in patients receiving cancer chemotherapy, treating bone
PT   marrow transplantation patients and for healing fractured bones  -
XX
PS   Claim 2; Page 56; 74pp; English.
XX
CC   AAC61879-93 encode secreted human proteins. The secreted proteins are
CC   useful in assays to determine their biological activities. The proteins
CC   can also be used as biomarkers to identify tissues or cell types which
CC   express the proteins. The polynucleotide molecules can be used as
CC   biomarkers for tissues or chromosomes and to elicit immune responses.
CC   The proteins and antibodies are useful in diagnosis and treatment of
CC   diseases associated with altered expression of these proteins. The
CC   proteins are also useful for prevention or treatment of platelet
CC   disorders, stem cell disorders, osteoporosis or osteoarthritis, burns,
CC   incisions, ulcers, periodontal diseases, central and peripheral nervous
CC   system diseases and neuropathies, for healing fractured bones and to
CC   induce cartilage and/or bone growth in cosmetic plastic surgery. The
CC   proteins are also useful for gut protection or regeneration, for the
CC   treatment of lung or liver fibrosis, for stimulating blood cell
CC   generation in patients receiving cancer chemotherapy and for treatment
CC   of bone marrow transplantation patients.
XX
SQ   Sequence 1839 BP; 509 A; 361 C; 424 G; 545 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 21;  Length 1839;
  Best Local Similarity   100.0%;  Pred. No. 90;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db       1494 GCAAACATCAAATAATATAA 1475
```

```
RESULT 26
AAV62752/c
ID   AAV62752 standard; cDNA; 1904 BP.
XX
AC   AAV62752;
XX
DT   15-FEB-1999  (first entry)
XX
DE   Human secreted protein clone fg949_3 cDNA.
XX
KW   Secreted protein; human; fg949_3; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             77..1096
FT                   /*tag=  a
FT   sig_peptide     128..166
FT                   /*tag=  b
FT   mat_peptide     167..1093
FT                   /*tag=  c
XX
PN   WO9846757-A2.
XX
PD   22-OCT-1998.
XX
PF   14-APR-1998;   98WO-US07999.
XX
PR   13-APR-1998;   98US-0059487.
PR   15-APR-1997;   97US-0843374.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  LaVallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1998-568731/48.
DR   P-PSDB; AAW74726.
XX
PT   New polynucleotide(s) encoding secreted human proteins - are derived
PT   from, e.g. human foetal brain or foetal kidney cDNA libraries,
PT   potentially useful as, e.g. vaccines or thrombolytic agents
XX
PS   Claim 28; Page 89-90; 120pp; English.
XX
CC   Full-length cDNA clone fg949_3 includes an open reading frame
CC   encoding a human secreted protein (see AAW74726).  It was isolated
CC   from a human adult brain cDNA library using methods which are
CC   selective for cDNAs encoding secreted proteins, or was identified
CC   as encoding a secreted or transmembrane protein on the basis of
CC   computer analysis of the amino acid sequence of the encoded protein.
CC   It shows some similarity to database sequences.  The invention
CC   provides polynucleotides (see AAV62746-55) from human foetal brain,
CC   adult testis, adult brain, adult kidney and foetal kidney (all
CC   deposited as composite clone ATCC 98404), which encode human
CC   secreted proteins (see AAW74720-29).  The polynucleotides and
CC   proteins are predicted to have biological activities which would
CC   make them suitable for treating, preventing or ameliorating medical
CC   conditions in humans and animals, although no supporting data are
CC   given.  Suggested activities include nutritional activity, immune
CC   stimulating (e.g. as vaccines) or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity,
CC   haemostatic and thrombolytic activity, receptor/ligand activity,
CC   antiinflammatory activity, cadherin/tumour invasion suppressor
CC   activity, and tumour inhibition activity.  The polynucleotides are
CC   also stated to be useful for gene therapy, and can be used in
CC   recombinant production of the polypeptidess.
XX
SQ   Sequence 1904 BP; 532 A; 381 C; 439 G; 551 T; 1 other;
```

```
 Query Match             0.2%;  Score 20;  DB 19;  Length 1904;
 Best Local Similarity   100.0%;  Pred. No. 90;
 Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db       1531 GCAAACATCAAATAATATAA 1512


RESULT 27
ABQ92055/c
ID   ABQ92055 standard; cDNA; 1904 BP.
XX
AC   ABQ92055;
XX
DT   04-OCT-2002  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 52.
XX
KW   Human; cytostatic; antirheumatic; antiarthritic; vulnerary; analgesic;
KW   antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
KW   neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;
KW   antiulcer; fungicide; antidiabetic; antiasthmatic; antiallergic;
KW   immunostimulant; antiparasitic; secreted protein; transmembrane protein;
KW   cytokine; cell proliferation; cell differentiation; autoimmune disease;
KW   stem cell; growth factor; nervous system disease; neuropathy;
KW   Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW   osteoporosis; severe combined immunodeficiency; SCID; infection;
KW   multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   US2002065394-A1.
XX
PD   30-MAY-2002.
XX
PF   22-DEC-2000; 2000US-0745763.
XX
PR   18-MAR-1998;   98US-0040963.
XX
PA   (JACO/) JACOBS K.
PA   (MCCO/) MCCOY J M.
PA   (LAVA/) LAVALLIE E R.
PA   (COLL/) COLLINS-RACIE L A.
PA   (EVAN/) EVANS C.
PA   (MERB/) MERBERG D.
PA   (TREA/) TREACY M.
PA   (SPAU/) SPAULDING V.
XX
PI   Jacobs K,  McCoy JM,  LaVallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M,  Spaulding V;
XX
DR   WPI; 2002-582343/62.
DR   P-PSDB; ABP61841.
XX
PT   Novel secreted or transmembrane protein and polynucleotide encoding the
PT   protein, useful for diagnosis and treatment of neurological disorders,
PT   cancer, autoimmune diseases, bone disorders and lung or liver fibrosis
PT   -
XX
PS   Claim 183; Page 191-192; 284pp; English.
XX
CC   The invention relates to human secreted or transmembrane protein (I),
CC   their fragments and is encoded by specific complementary deoxyribonucleic
CC   acid (cDNA) inserts (II), where the protein is substantially free from
CC   other mammalian proteins. (I) are useful for preventing, treating or
CC   ameliorating a medical condition, especially immunological treatment or
CC   prevention of tumours. (I) exhibits activity relating to angiogenesis,
CC   cytokine, cell proliferation, cell differentiation, antiinflammatory,
CC   stem cell growth factor activity and activin or inhibin-related
CC   activities. (I) can be used to manipulate stem cells in culture to give
CC   rise to neuroepithelial cells that can be used to augment or replace
CC   cells damaged by illness, autoimmune disease, accidental damage or
CC   genetic disorders. (I) induces the proliferation of neural cells and
CC   regeneration of nerve and brain tissue and is useful for the treatment of
CC   central and peripheral nervous system diseases and neuropathies, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis. (I) is involved in chemotactic or chemokinetic
CC   activity, regulation of haematopoiesis and is useful for treating myeloid
CC   or lymphoid cell disorders, platelet disorders such as thrombocytopaenia
CC   and for regeneration of bone, cartilage, tendon, ligament and/or nerve
CC   tissue growth and in tissue repair, healing of burns, incisions, ulcers,
CC   for treating osteoporosis, osteoarthritis, bone degenerative disorders or
CC   periodontal disease. (I) is also useful for gut protection or
CC   regeneration and treatment of lung or liver fibrosis, reperfusion injury
CC   in various tissues, various immune deficiencies and disorders including
CC   severe combined immunodeficiency (SCID), bacterial or fungal infections,
CC   autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,
CC   diabetes mellitus, myasthenia gravis, allergic reactions and conditions,
CC   such as asthma or other respiratory problems. (II) is useful to express
CC   recombinant protein, as markers for tissues in which the corresponding
CC   protein is preferentially expressed and in gene therapy. The present
CC   sequence is that of a polynucleotide of the invention.
XX
SQ   Sequence 1904 BP; 532 A; 381 C; 439 G; 551 T; 1 other;

  Query Match             0.2%;  Score 20;  DB 24;  Length 1904;
  Best Local Similarity   100.0%;  Pred. No. 90;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db       1531 GCAAACATCAAATAATATAA 1512


RESULT 28
ABQ91913/c
ID   ABQ91913 standard; cDNA; 1924 BP.
XX
AC   ABQ91913;
XX
DT   30-SEP-2002  (first entry)
XX
DE   Human NF-kB activating gene SEQ ID NO 4.
XX
KW   Human; NF-kB; nuclear factor kappa.B; mouse; antiinflammatory;
KW   immunomodulator; cytostatic; antiinfective; osteopathic; nootropic;
KW   neuroprotective; anti-HIV; autoimmune disease; cancer; infection;
KW   bone disease; AIDS; neurodegenerative disease; ischaemic disorder;
KW   gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200253737-A1.
XX
PD   11-JUL-2002.
XX
PF   25-DEC-2001; 2001WO-JP11389.
XX
PR   28-DEC-2000; 2000JP-0402288.
PR   26-MAR-2001; 2001JP-0088912.
PR   24-AUG-2001; 2001JP-0254018.
XX
PA   (ASAH ) ASAHI KASEI KOGYO KK.
XX
PI   Matsuda A,  Honda G,  Muramatsu S,  Nagano Y;
XX
DR   WPI; 2002-583617/62.
DR   P-PSDB; ABP61425.
XX
PT   NF-approximatelykB activating gene and expressed protein, applicable in
PT   diagnosis and screening inhibitors or promoters to control excessive
PT   activation or inhibition for treating e.g. inflammations, autoimmune
PT   diseases and cancer  -
XX
PS   Claim 4; Page 107-111; 841pp; Japanese.
```

```
XX
CC   The invention relates to a purified protein (I), comprising one of 90
CC   fully defined sequences (ABP61424-ABP61513) or a protein based on any of
CC   the sequences but with some amino acids deleted, substituted or added and
CC   with a NF-kB (nuclear factor kappa B) activating effect. The protein and
CC   encoding gene (ABQ91912-ABQ92001) are useful in diagnosis and screening
CC   inhibitors or promoters to control excessive activation or inhibition
CC   and for treating e.g. inflammations, autoimmune diseases, cancers,
CC   infections, bone diseases, AIDS, neurodegenerative diseases or ischaemic
CC   disorders.
XX
SQ   Sequence 1924 BP; 514 A; 393 C; 458 G; 559 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 24;  Length 1924;
  Best Local Similarity   100.0%;  Pred. No. 90;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db       1569 GCAAACATCAAATAATATAA 1550


RESULT 29
AAF30021/c
ID   AAF30021 standard; cDNA; 1925 BP.
XX
AC   AAF30021;
XX
DT   23-APR-2001  (first entry)
XX
DE   Human hydrophobic domain-containing protein cDNA clone HP03342.
XX
KW   Human; hydrophobic domain; antiinflammatory; immunosuppressive;
KW   immunostimulant; vulnerary; antiulcer; haemostatic; cytostatic;
KW   hepatotropic; thrombolytic; antitumour; antiviral; antibacterial;
KW   antifungal; gene therapy; diagnosis; membrane protein;
KW   protein secretion; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             115..1134
FT                   /*tag= a
FT                   /note= "the open reading frame is also
FT                          specifically claimed, in Claim 3"
XX
PN   WO200100824-A2.
XX
PD   04-JAN-2001.
XX
PF   16-JUN-2000; 2000WO-JP03944.
XX
PR   24-JUN-1999;   99JP-0178065.
XX
PA   (SAGA ) SAGAMI CHEM RES CENT.
PA   (PROT-) PROTEGENE INC.
XX
PI   Kato S,  Kimura T;
XX
DR   WPI; 2001-123008/13.
DR   P-PSDB; AAB20090.
XX
PT   New proteins with hydrophobic domains, useful for treating autoimmune
PT   diseases, inflammatory diseases, wounds, burns, incisions and ulcers  -
XX
PS   Claim 4; Page 106-110; 126pp; English.
XX
CC   The present sequence is that of cDNA clone HP03342 encoding a
CC   human protein (see AAB20090) having a predicted N-terminal
CC   secretory signal and a transmembrane domain.  The clone was
CC   obtained from a human umbilical cord blood cDNA library.  The
CC   predicted protein shows 35% homology to the Caenorhabditis elegans
CC   hypothetical protein CET06D8.9.  The invention provides human
CC   proteins (see AAB20089-96) having hydrophobic domains indicative
CC   of secretory signals and membrane domains, and nucleic acids (see
CC   AAF30020-27) encoding them.  These are expected to have cell
CC   proliferation, cell differentiation, immunostimulant or
CC   immunosuppressor, haematopoiesis regulating, tissue growth,
CC   activin, inhibin, chemotactic, chemokinetic, haemostatic,
CC   thrombolytic, receptor/ligand, antiinflammatory or antitumour
CC   activities (no evidence provided).  Expression vectors, eukaryotic
CC   host cells and antibodies are also provided.  The human cDNAs of
CC   the invention are used as probes for genetic diagmosis and gene
CC   sources for gene therapy, or are used in the recombinant production
CC   of the proteins.  Cells expressing the secretory or membrane
CC   proteins are utilized for detection of corresponding receptors of
CC   ligands, and in candidate drug screening.
XX
SQ   Sequence 1925 BP; 515 A; 393 C; 457 G; 560 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 22;  Length 1925;
  Best Local Similarity   100.0%;  Pred. No. 90;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db       1569 GCAAACATCAAATAATATAA 1550


RESULT 30
AAF93751/c
ID   AAF93751 standard; cDNA; 1927 BP.
XX
AC   AAF93751;
XX
DT   23-MAY-2001  (first entry)
XX
DE   Human cDNA encoding a membrane or secretory protein clone PSEC0019.
XX
KW   Human; secretory protein; membrane protein; vaccine; gene therapy;
KW   rheumatoid arthritis; diabetes; ss.
XX
OS   Homo sapiens.
XX
PN   EP1067182-A2.
XX
PD   10-JAN-2001.
XX
PF   07-JUL-2000; 2000EP-0114090.
XX
PR   08-JUL-1999;   99JP-0194179.
PR   11-JAN-2000; 2000JP-0118775.
PR   02-MAY-2000; 2000JP-0183766.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Kawai Y,  Sugiyama T,  Hayashi K;
XX
DR   WPI; 2001-093989/11.
DR   P-PSDB; AAB88324.
XX
PT   Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT   gene therapy or as candidate target molecules in drug development -
XX
PS   Claim 1; SEQ ID 15; 609pp + CD ROM; English.
XX
CC   This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC   which encode human secretory or membrane proteins represented by
CC   AAB88317 - AAB88419. Included in the invention are primers
CC   AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC   cDNA sequences of the invention. The invention also includes methods for
CC   the production of antibodies directed against the proteins, and cDNA
CC   sequences, which can be used in vaccines. The polynucleotide sequences
CC   can be used in gene therapy. The polynucleotide sequences and the
CC   proteins they encode may be used in the prevention, treatment and
CC   diagnosis of diseases associated with inappropriate secretory
```

```
CC  protein/membrane protein expression. The nucleic acids and complementary
CC  sequences may also be used as DNA probes in diagnostic assays
CC  (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC  presence of similar nucleic acid sequences in samples. They may also be
CC  used to study the expression and function of secretory proteins/membrane
CC  polypeptides and their role in metabolism. The polypeptides may be used
CC  as antigens in the production of antibodies against them and in assays to
CC  identify modulators (agonists and antagonists) of expression and
CC  activity. The antibodies and antagonists may also be used as therapeutic
CC  agents to down regulate expression and activity. The antibodies may also
CC  be used as diagnostic agents for detecting the presence of the
CC  polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC  (ELISA). Examples of diseases which may be treated include rheumatoid
CC  arthritis and diabetes.
XX
SQ  Sequence 1927 BP; 516 A; 395 C; 458 G; 558 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 22;  Length 1927;
  Best Local Similarity   100.0%;  Pred. No. 90;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db       1572 GCAAACATCAAATAATATAA 1553


RESULT 31
AAA96358/c
ID  AAA96358 standard; cDNA; 1977 BP.
XX
AC  AAA96358;
XX
DT  08-FEB-2001  (first entry)
XX
DE  cDNA encoding a novel polypeptide designated PRO4499.
XX
KW  Secreted protein; transmembrane protein; PRO1484; PRO4334; PRO1122;
KW  PRO1889; PRO1890; PRO1887; PRO1785; PRO4353; PRO4357; PRO4405; PRO4356;
KW  PRO4352; PRO4380; PRO4354; PRO4408; PRO5737; PRO4425; PRO5990; PRO6030;
KW  PRO4424; PRO4422; PRO4430; PRO4499; tumour; obesity; diabetes;
KW  insulinemia; kidney disorder; Bergers disease; nephropathy;
KW  Schonlein-Henoch purpura; celiac disease; dermatitis herpetiformis;
KW  Crohns disease; ss.
XX
OS  Homo sapiens.
XX
FH  Key             Location/Qualifiers
FT  CDS             185..1204
FT                  /*tag= a
FT  sig_peptide     185..274
FT                  /*tag= b
XX
PN  WO200056889-A2.
XX
PD  28-SEP-2000.
XX
PF  01-MAR-2000; 2000WO-US05601.
XX
PR  23-MAR-1999;  99US-0125774.
PR  23-MAR-1999;  99US-0125778.
PR  24-MAR-1999;  99US-0125826.
PR  31-MAR-1999;  99US-0127035.
PR  05-APR-1999;  99US-0127706.
PR  21-APR-1999;  99US-0130359.
PR  27-APR-1999;  99US-0131270.
PR  27-APR-1999;  99US-0131272.
PR  27-APR-1999;  99US-0131291.
PR  04-MAY-1999;  99US-0132371.
PR  04-MAY-1999;  99US-0132379.
PR  04-MAY-1999;  99US-0132383.
PR  25-MAY-1999;  99US-0135750.
PR  08-JUN-1999;  99US-0138166.
PR  20-JUL-1999;  99US-0144791.
PR  03-AUG-1999;  99US-0146970.
PR  09-DEC-1999;  99US-0170262.
XX
PA  (GETH ) GENENTECH INC.
XX
PI  Desnoyers L,  Eaton DL,  Goddard A,  Godowski PJ,  Gurney AL,  Pan J;
PI  Stewart TA,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR  WPI; 2000-628263/60.
DR  P-PSDB; AAB18931.
XX
PT  Novel secreted and transmembrane polypeptides useful for diagnosing
PT  tumour in a mammal, for identifying agonists and antagonists of the
PT  polypeptide and for therapeutic use
XX
PS  Claim 2; Fig 45; 222pp; English.
XX
CC  The present sequence encodes a secreted or transmembrane polypeptide.
CC  The specification describes polypeptides designated PRO1484, PRO4334,
CC  PRO1122, PRO1889, PRO1890, PRO1887, PRO1785, PRO4353, PRO4357, PRO4405,
CC  PRO4356, PRO4352, PRO4380, PRO4354, PRO4408, PRO5737, PRO4425, PRO5990,
CC  PRO6030, PRO4424, PRO4422, PRO4430 and PRO4499. PRO1889 polypeptide is
CC  useful for diagnosing tumour in a mammal. The polypeptides, their
CC  agonists and antagonists are useful treating a condition associated with
CC  expression or activity of the polypeptide. Conditions treated include
CC  obesity, diabetes or hyper-or hypo-insulinemia. The polypeptides are
CC  capable of inducing proliferation of mammalian kidney mesangial cells
CC  and are therefore useful-for treating kidney disorders associated with
CC  decreased mesangial cell function such as Bergers disease or other
CC  nephropathies associated with Schonlein-Henoch purpura, celiac disease,
CC  dermatitis herpetiformis or Crohns disease. The nucleic acids may be used
CC  to generate transgenic animals for use in development and screening of
CC  therapeutically useful reagents and also for chromosome identification
CC  and tissue typing.
XX
SQ  Sequence 1977 BP; 513 A; 412 C; 482 G; 570 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 21;  Length 1977;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db       1639 GCAAACATCAAATAATATAA 1620


RESULT 32
AAS46176/c
ID  AAS46176 standard; cDNA; 1977 BP.
XX
AC  AAS46176;
XX
DT  18-DEC-2001  (first entry)
XX
DE  Human DNA encoding PRO polypeptide sequence #252.
XX
KW  PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
KW  dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW  blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW  adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
KW  PCR primer.
XX
OS  Homo sapiens.
XX
PN  WO200168848-A2.
XX
PD  20-SEP-2001.
XX
PF  28-FEB-2001; 2001WO-US06520.
XX
PR  01-MAR-2000; 2000WO-US05601.
PR  02-MAR-2000; 2000WO-US05841.
PR  03-MAR-2000; 2000US-187202P.
```

PR 06-MAR-2000; 2000US-186968P.
PR 14-MAR-2000; 2000US-189320P.
PR 14-MAR-2000; 2000US-189328P.
PR 15-MAR-2000; 2000WO-US06884.
PR 21-MAR-2000; 2000US-190828P.
PR 21-MAR-2000; 2000US-191007P.
PR 21-MAR-2000; 2000US-191048P.
PR 21-MAR-2000; 2000US-191314P.
PR 28-MAR-2000; 2000US-192655P.
PR 29-MAR-2000; 2000US-193032P.
PR 29-MAR-2000; 2000US-193053P.
PR 30-MAR-2000; 2000WO-US08439.
PR 04-APR-2000; 2000US-194449P.
PR 04-APR-2000; 2000US-194647P.
PR 11-APR-2000; 2000US-195975P.
PR 11-APR-2000; 2000US-196000P.
PR 11-APR-2000; 2000US-196187P.
PR 11-APR-2000; 2000US-196690P.
PR 11-APR-2000; 2000US-196820P.
PR 18-APR-2000; 2000US-198121P.
PR 18-APR-2000; 2000US-198585P.
PR 25-APR-2000; 2000US-199397P.
PR 25-APR-2000; 2000US-199550P.
PR 25-APR-2000; 2000US-199654P.
PR 03-MAY-2000; 2000US-201516P.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 05-JUN-2000; 2000US-209832P.
PR 28-JUL-2000; 2000WO-US20710.
PR 22-AUG-2000; 2000US-0644848.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
DR WPI; 2001-602746/68.
DR P-PSDB; AAU29275.
XX
PT Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT presence of tumours, such as prostate and breast tumours, in mammals and
PT to screen for modulators of the compounds -
XX
PS Claim 2; Fig 503; 774pp; English.
XX
CC Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR
CC primers for PRO polypeptides of the invention. The sequences of the
CC invention can be used to detect the presence of a tumour in a mammal by
CC comparing the level of expression of a PRO polypeptide in a test sample
CC of cells from the animal and a control sample of normal cells, whereby a
CC higher level of expression in the test sample indicates the presence of a
CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,
CC pigs, goats and rabbits but are preferably human. The polypeptides can be
CC used to stimulate tumour necrosis factor (TNF) alpha release from human
CC blood, when contacted with it. A specific polypeptide can be used to
CC stimulate the proliferation or differentiation of chondrocyte cells. The
CC PRO proteins can be used to determine the presence of tumours and also
CC susceptibility to tumour development, particularly adrenal, lung, colon,
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC can be used for genetic analysis of individuals with genetic disorders.
XX
SQ Sequence 1977 BP; 513 A; 412 C; 482 G; 570 T; 0 other;

Query Match 0.2%; Score 20; DB 22; Length 1977;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4861 GCAAACATCAAATAATATAA 4880
||||||||||||||||||||
Db 1639 GCAAACATCAAATAATATAA 1620

RESULT 33
ABL95723/c
ID ABL95723 standard; cDNA; 1977 BP.
XX
AC ABL95723;
XX
DT 19-JUL-2002 (first entry)
XX
DE Human angiogenesis related cDNA PRO4499 SEQ ID NO: 325.
XX
KW Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer;
KW atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder;
KW cardiant; cytostatic; antiangiogenic; hypotensive; vulnerary;
KW antiarteriosclerotic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200208284-A2.
XX
PD 31-JAN-2002.
XX
PF 09-JUL-2001; 2001WO-US21735.
XX
PR 20-JUL-2000; 2000US-219556P.
PR 25-JUL-2000; 2000US-220624P.
PR 25-JUL-2000; 2000US-220664P.
PR 28-JUL-2000; 2000WO-US20710.
PR 02-AUG-2000; 2000US-222695P.
PR 17-AUG-2000; 2000US-0643657.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 07-SEP-2000; 2000US-230978P.
PR 15-SEP-2000; 2000US-000000P.
PR 18-SEP-2000; 2000US-0664610.
PR 18-SEP-2000; 2000US-0665350.
PR 24-OCT-2000; 2000US-242922P.
PR 08-NOV-2000; 2000US-0709238.
PR 08-NOV-2000; 2000WO-US30952.
PR 10-NOV-2000; 2000WO-US30873.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000US-0747259.
PR 20-DEC-2000; 2000WO-US34956.
PR 22-JAN-2001; 2001US-0767609.
PR 28-FEB-2001; 2001US-0796498.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-MAR-2001; 2001WO-US06666.
PR 09-MAR-2001; 2001US-0802706.
PR 14-MAR-2001; 2001US-0808689.
PR 22-MAR-2001; 2001US-0816744.
PR 05-APR-2001; 2001US-0828366.
PR 10-MAY-2001; 2001US-0854208.
PR 10-MAY-2001; 2001US-0854280.
PR 25-MAY-2001; 2001US-0866028.
PR 25-MAY-2001; 2001US-0866034.
PR 25-MAY-2001; 2001WO-US17092.
PR 30-MAY-2001; 2001US-0870574.
PR 30-MAY-2001; 2001WO-US17443.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 28-JUN-2001; 2001WO-US00000.
XX
PA (GETH) GENENTECH INC.
PA (BAKE/) BAKER K P.
PA (FERR/) FERRARA N.
PA (GERB/) GERBER H.
PA (GERR/) GERRITSEN M E.
PA (GODD/) GODDARD A.

PA (GODO/) GODOWSKI P J.
PA (GURN/) GURNEY A L.
PA (HILL/) HILLAN K J.
PA (MARS/) MARSTERS S A.
PA (PANJ/) PAN J.
PA (PAON/) PAONI N F.
PA (STEP/) STEPHAN J F.
PA (WATA/) WATANABE C K.
PA (WILL/) WILLIAMS P M.
PA (WOOD/) WOOD W I.
XX
PI Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;
PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;
PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
XX
DR WPI; 2002-171999/22.
DR P-PSDB; ABB95585.
XX
PT One hundred and eighty seven nucleic acids encoding PRO polypeptides,
PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
PT infarction), endothelial or angiogenic disorders in a mammal -
XX
PS Claim 1; Fig 325; 567pp; English.
XX
CC The present invention provides the protein and coding sequences of human
CC PRO proteins. These are useful for treating or diagnosing a
CC cardiovascular, endothelial or angiogenic disorder, including cardiac
CC hypertrophy, trauma, cancer, age-related macular degeneration,
CC atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis,
CC angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour
CC angiogenesis (such as breast carcinoma and liver carcinoma) and wound
CC healing. The present sequence is a coding sequence of the invention.
XX
SQ Sequence 1977 BP; 513 A; 412 C; 482 G; 570 T; 0 other;

Query Match 0.2%; Score 20; DB 24; Length 1977;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4861 GCAAACATCAAATAATATAA 4880
||||||||||||||||||||
Db 1639 GCAAACATCAAATAATATAA 1620

RESULT 34
ABK69983/c
ID ABK69983 standard; DNA; 1977 BP.
XX
AC ABK69983;
XX
DT 15-JUL-2002 (first entry)
XX
DE cDNA encoding human Pro peptide #23.
XX
KW Human; ss; gene; PRO; secreted protein; transmembrane protein;
KW genetic disorder; tumour; cancer.
XX
OS Homo sapiens.
XX
PN WO200224888-A2.
XX
PD 28-MAR-2002.
XX
PF 29-AUG-2001; 2001WO-US27099.
XX
PR 01-SEP-2000; 2000US-229896P.
PR 05-SEP-2000; 2000US-230621P.
PR 22-SEP-2000; 2000US-235147P.
PR 10-NOV-2000; 2000WO-US30873.
PR 12-JAN-2001; 2001US-261878P.
PR 16-JAN-2001; 2001US-261910P.
PR 16-JAN-2001; 2001US-261939P.
PR 16-JAN-2001; 2001US-262150P.
PR 25-JAN-2001; 2001US-264395P.
PR 02-FEB-2001; 2001US-266421P.
PR 09-FEB-2001; 2001US-267623P.
PR 28-FEB-2001; 2001WO-US06520.
PR 09-MAR-2001; 2001US-274399P.
PR 03-APR-2001; 2001US-280982P.
PR 04-APR-2001; 2001US-282129P.
PR 04-APR-2001; 2001US-282199P.
PR 09-MAY-2001; 2001US-290589P.
PR 25-MAY-2001; 2001WO-US17092.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Eaton DL, Filvaroff E, Goddard A, Grimaldi JC;
PI Gurney AL, Smith V, Stephan J, Watanabe CK, Wood WI, Zhang Z;
PI Fong S;
XX
DR WPI; 2002-362426/39.
DR P-PSDB; ABG34052.
XX
PT New PRO polypeptides and polynucleotides encoding the polypeptides,
PT useful in gene therapy, chromosome identification, tissue typing, or
PT for genetic analysis of individuals with genetic disorders -
XX
PS Claim 2; Figure 45; 218pp; English.
XX
CC This invention relates to the cDNA and protein sequences of novel
CC secreted and transmembrane polypeptides PRO polypeptides. The
CC invention also comprises a method for producing the proteins of the
CC invention by recombinant means and antibodies specific for the protein
CC of the invention. The antibody may be used for detecting the PRO
CC proteins of the invention and may be used to modify their activity.
CC polynucleotides may be used as hybridisation probes for a cDNA library
CC to isolate the full-length PRO cDNA or to isolate other cDNAs, to
CC construct hybridisation probes for mapping the gene which encodes that
CC PRO and for genetic analysis of individuals with genetic disorders, in
CC assays to identify other proteins or molecules involved in binding
CC reaction, to generate transgenic animals or knock-out animals which in
CC turn are useful in the development and screening of therapeutically
CC useful reagents, for chromosome identification, and tissue typing. The
CC PRO polypeptides are useful in gene therapy, and as molecular weight
CC markers for protein electrophoresis purposes. The sequences may
CC also be used to detect overexpression on PRO polypeptides in cancerous
CC tumours and for screening for differentially expressed genes using
CC microarray technology. The present sequence represents a cDNA encoding
CC a human PRO protein of the invention.
XX
SQ Sequence 1977 BP; 513 A; 412 C; 482 G; 570 T; 0 other;

Query Match 0.2%; Score 20; DB 24; Length 1977;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4861 GCAAACATCAAATAATATAA 4880
||||||||||||||||||||
Db 1639 GCAAACATCAAATAATATAA 1620

RESULT 35
ABL88234/c
ID ABL88234 standard; cDNA; 1977 BP.
XX
AC ABL88234;
XX
DT 16-MAY-2002 (first entry)
XX
DE Human PRO4499 cDNA sequence SEQ ID NO:325.
XX
KW Human; angiogenesis; cardiant; cytostatic; antiangiogenic; hypotensive;

KW vulnerary; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma;
KW gene therapy; cardiovascular disorder; endothelial disorder; cancer;
KW angiogenic disorder; cardiac hypertrophy; atherosclerosis; hypertension;
KW age-related macular degeneration; arterial restenosis; angina;
KW rheumatoid arthritis; myocardial infarction; thrombophlebitis;
KW lymphangitis; tumour angiogenesis; breast carcinoma; liver carcinoma;
KW wound healing; chromosome mapping; gene mapping; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200200690-A2.
XX
PD 03-JAN-2002.
XX
PF 20-JUN-2001; 2001WO-US19692.
XX
PR 23-JUN-2000; 2000US-213637P.
PR 20-JUL-2000; 2000US-219556P.
PR 25-JUL-2000; 2000US-220624P.
PR 25-JUL-2000; 2000US-220664P.
PR 28-JUL-2000; 2000WO-US20710.
PR 02-AUG-2000; 2000US-222695P.
PR 17-AUG-2000; 2000US-0643657.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 07-SEP-2000; 2000US-230978P.
PR 18-SEP-2000; 2000US-0664610.
PR 18-SEP-2000; 2000US-0665350.
PR 24-OCT-2000; 2000US-242922P.
PR 08-NOV-2000; 2000US-0709238.
PR 08-NOV-2000; 2000WO-US30952.
PR 10-NOV-2000; 2000WO-US30873.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000US-0747259.
PR 20-DEC-2000; 2000WO-US34956.
PR 22-JAN-2001; 2001US-0767609.
PR 28-FEB-2001; 2001US-0796498.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-MAR-2001; 2001WO-US06666.
PR 09-MAR-2001; 2001US-0802706.
PR 14-MAR-2001; 2001US-0808689.
PR 22-MAR-2001; 2001US-0816744.
PR 05-APR-2001; 2001US-0828366.
PR 10-MAY-2001; 2001US-0854208.
PR 10-MAY-2001; 2001US-0854280.
PR 25-MAY-2001; 2001US-0866028.
PR 25-MAY-2001; 2001US-0866034.
PR 25-MAY-2001; 2001WO-US17092.
PR 30-MAY-2001; 2001US-0870574.
PR 30-MAY-2001; 2001WO-US17443.
PR 01-JUN-2001; 2001WO-US17800.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;
PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;
PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
XX
DR WPI; 2002-090516/12.
DR P-PSDB; ABB84979.
XX
PT One hundred and eighty seven nucleic acids encoding PRO polypeptides,
PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
PT infarction), endothelial or angiogenic disorders in a mammal -
XX
PS Claim 2; Fig 325; 565pp; English.
XX
CC ABL88072 to ABL88258 encode the PRO proteins given in ABB84817 to
CC ABB85003. The PRO proteins and polynucleotides have cardiant, cytostatic,
CC antiangiogenic, hypotensive, vulnerary and antiarteriosclerotic
CC activities, and can be used in gene therapy. The PRO polynucleotides,
CC proteins, agonists and antagonists are useful for treating or diagnosing
CC a cardiovascular, endothelial or angiogenic disorder in a mammal,
CC e.g. cardiac hypertrophy, trauma, cancer, age-related macular
CC degeneration, atherosclerosis, hypertension, arterial restenosis,
CC rheumatoid arthritis, angina, myocardial infarctions, thrombophlebitis,
CC lymphangitis, tumour angiogenesis (such as breast carcinoma and liver
CC carcinoma) and wound healing. The PRO polynucleotides have applications
CC in molecular biology, including use as hybridisation probes, and in
CC chromosome and gene mapping. ABL88259 to ABL88267 represent primers and
CC probes used in the exemplification of the present invention.
XX
SQ Sequence 1977 BP; 513 A; 412 C; 482 G; 570 T; 0 other;

Query Match 0.2%; Score 20; DB 24; Length 1977;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4861 GCAAACATCAAATAATATAA 4880
||||||||||||||||||||
Db 1639 GCAAACATCAAATAATATAA 1620

RESULT 36
ACA06124/c
ID ACA06124 standard; cDNA; 1977 BP.
XX
AC ACA06124;
XX
DT 02-JUN-2003 (first entry)
XX
DE cDNA encoding human PRO polypeptide #23.
XX
KW Human; PRO polypeptide; secreted and transmembrane protein; cancer;
KW non-insulin dependent diabetes mellitus; septic shock; stroke;
KW rheumatoid arthritis; graft-versus-host disease; cardiac ischaemia;
KW psoriasis; inflammatory bowel disease; asthma; antidiabetic;
KW cytostatic; immunosuppressive; antirheumatic; antiarthritic;
KW cerebroprotective; vasotropic; antipsoriatic; antiinflammatory;
KW antiasthmatic; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN US2003008348-A1.
XX
PD 09-JAN-2003.
XX
PF 26-DEC-2001; 2001US-0035855.
XX
PR 14-MAY-1999; 99WO-US10733.
PR 02-DEC-1999; 99WO-US28551.
PR 22-DEC-1999; 99WO-US30720.
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 22-MAY-2000; 2000WO-US14042.
PR 02-JUN-2000; 2000WO-US15264.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 01-DEC-2000; 2000WO-US32678.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 15-MAY-1998; 98US-085579P.
PR 15-DEC-1998; 98US-112514P.
PR 22-DEC-1998; 98US-113300P.
PR 23-DEC-1998; 98US-113430P.
PR 23-DEC-1998; 98US-113605P.
PR 22-JAN-1999; 99US-116843P.
PR 23-MAR-1999; 99US-125774P.
PR 23-MAR-1999; 99US-125778P.
PR 24-MAR-1999; 99US-125826P.
PR 31-MAR-1999; 99US-127035P.
PR 05-APR-1999; 99US-127706P.
PR 13-APR-1999; 99US-129122P.
PR 21-APR-1999; 99US-130359P.

```
PR  25-MAY-1999;   99US-135750P.
PR  08-JUN-1999;   99US-138166P.
PR  20-JUL-1999;   99US-144791P.
PR  03-AUG-1999;   99US-146970P.
PR  29-OCT-1999;   99US-162506P.
PR  16-AUG-2001; 2001US-0931836.
XX
PA  (GETH ) GENENTECH INC.
XX
PI  Desnoyers L,  Eaton DL,  Goddard A,  Godowski PJ,  Gurney AL,  Pan J;
PI  Stewart TA,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR  WPI; 2003-341326/32.
DR  P-PSDB; ABU69102.
XX
PT  New PRO polypeptides and nucleic acid molecules, useful for diagnosing
PT  or treating diabetes mellitus, cancers, septic shock, inflammatory
PT  bowel disease or asthma, or in gene therapy, chromosome identification
PT  or tissue typing -
XX
PS  Claim 2; Fig 45; 196pp; English.
XX
CC  The present invention relates to the isolation of novel human PRO
CC  polypeptides, and the polynucleotide sequences encoding them. The
CC  PRO polypeptides are secreted and transmembrane proteins. The PRO
CC  polypeptides and polynucleotides are useful in diagnosing or
CC  treating non-insulin dependent diabetes mellitus, cancers, septic
CC  shock, rheumatoid arthritis, graft-versus-host disease, stroke, cardiac
CC  ischaemia, psoriasis, inflammatory bowel disease or asthma. The PRO
CC  polynucleotide sequences may be used as hybridisation probes in
CC  chromosome and gene mapping, or in generating antisense RNA and DNA.
CC  They are also useful in preparing PRO polypeptides, in assays to
CC  identify other proteins or molecules involved in binding reaction, to
CC  generate transgenic animals or knockout animals, which in turn are
CC  useful in the development and screening of therapeutically useful
CC  reagents, for chromosome identification, and tissue typing. The PRO
CC  polypeptides and nucleic acid molecules are also useful in gene
CC  therapy, and as molecular weight markers for protein electrophoresis
CC  purposes. Anti-PRO antibodies may be used in diagnostic assays for
CC  PRO polypeptides, or for the affinity purification of the polypeptides
CC  from recombinant cell culture or natural sources. The present sequence
CC  represents a cDNA encoding a human PRO polypeptide of the invention.
XX
SQ  Sequence 1977 BP; 513 A; 412 C; 482 G; 570 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 25;  Length 1977;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db       1639 GCAAACATCAAATAATATAA 1620


RESULT 37
ACA06181/c
ID  ACA06181 standard; cDNA; 1977 BP.
XX
AC  ACA06181;
XX
DT  02-JUN-2003  (first entry)
XX
DE  cDNA encoding human PRO polypeptide #23.
XX
KW  Human; secreted and transmembrane protein; bone disorder; obesity;
KW  cartilage disorder; sports injury; arthritis; diabetes mellitus;
KW  hypo-insulinaemia; obesity; hyper-insulinaemia; thalassaemia;
KW  haemoglobin-associated disorder; kidney disorder; Berger disease;
KW  mesangial cell function; nephropathy; Schonlein-Henoch purpura;
KW  celiac disease; dermatitis herpetiformis; Crohn's disease;
KW  anorectic; antiarthritic; antidiabetic; antianaemic; nephrotropic;
KW  antiinflammatory; gene therapy; gene; ss.
XX
OS  Homo sapiens.
XX
PN  US2003032061-A1.
XX
PD  13-FEB-2003.
XX
PF  26-DEC-2001; 2001US-0036214.
XX
PR  14-MAY-1999;   99WO-US10733.
PR  02-DEC-1999;   99WO-US28551.
PR  22-DEC-1999;   99WO-US30720.
PR  01-MAR-2000; 2000WO-US05601.
PR  02-MAR-2000; 2000WO-US05841.
PR  22-MAY-2000; 2000WO-US14042.
PR  02-JUN-2000; 2000WO-US15264.
PR  23-AUG-2000; 2000WO-US23522.
PR  24-AUG-2000; 2000WO-US23328.
PR  01-DEC-2000; 2000WO-US32678.
PR  20-DEC-2000; 2000WO-US34956.
PR  28-FEB-2001; 2001WO-US06520.
PR  01-JUN-2001; 2001WO-US17800.
PR  20-JUN-2001; 2001WO-US19692.
PR  29-JUN-2001; 2001WO-US21066.
PR  09-JUL-2001; 2001WO-US21735.
PR  15-MAY-1998;   98US-085579P.
PR  15-DEC-1998;   98US-112514P.
PR  22-DEC-1998;   98US-113300P.
PR  23-DEC-1998;   98US-113430P.
PR  23-DEC-1998;   98US-113605P.
PR  23-DEC-1998;   98US-113621P.
PR  23-DEC-1998;   98US-114140P.
PR  12-JAN-1999;   99US-115552P.
PR  22-JAN-1999;   99US-116843P.
PR  23-MAR-1999;   99US-125774P.
PR  23-MAR-1999;   99US-125778P.
PR  24-MAR-1999;   99US-125826P.
PR  31-MAR-1999;   99US-127035P.
PR  05-APR-1999;   99US-127706P.
PR  13-APR-1999;   99US-129122P.
PR  21-APR-1999;   99US-130359P.
PR  27-APR-1999;   99US-131270P.
PR  27-APR-1999;   99US-131272P.
PR  27-APR-1999;   99US-131291P.
PR  04-MAY-1999;   99US-132371P.
PR  04-MAY-1999;   99US-132379P.
PR  04-MAY-1999;   99US-132383P.
PR  25-MAY-1999;   99US-135750P.
PR  08-JUN-1999;   99US-138166P.
PR  20-JUL-1999;   99US-144791P.
PR  03-AUG-1999;   99US-146970P.
PR  29-OCT-1999;   99US-162506P.
PR  16-AUG-2001; 2001US-0931836.
XX
PA  (GETH ) GENENTECH INC.
XX
PI  Desnoyers L,  Eaton DL,  Goddard A,  Godowski PJ,  Gurney AL,  Pan J;
PI  Stewart TA,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR  WPI; 2003-341962/32.
DR  P-PSDB; ABU69125.
XX
PT  Novel isolated PRO polypeptides e.g., PRO4334, PRO1122, PRO1889,
PT  PRO1890, PRO1887, PRO1785, PRO4353, useful for treating sports
PT  injuries, arthritis, diabetes, obesity, hyper- or hypo-insulinaemia -
XX
PS  Claim 2; Fig 45; 194pp; English.
XX
CC  The present invention relates to the isolation of novel human PRO
CC  polypeptides, and the polynucleotide sequences encoding them. The
CC  PRO polypeptides are secreted and transmembrane proteins. The PRO
CC  polypeptides and polynucleotides are useful in diagnosing or
CC  treating various bone and/or cartilage disorders (e.g. sports
CC  injuries, arthritis), various insulin deficient states (e.g. diabetes
```

```
CC   mellitus, hypo-insulinaemia), obesity, hyper-insulinaemia,
CC   haemoglobin-associated disorders (e.g. thalassaemias), kidney disorders
CC   associated with decreased mesangial cell function (e.g. Berger disease),
CC   or other nephropathies associated with Schonlein-Henoch purpura,
CC   celiac disease, dermatitis herpetiformis or Crohn's disease. The PRO
CC   polynucleotide sequences may be used as hybridisation probes in
CC   chromosome and gene mapping, or in generating antisense RNA and DNA.
CC   They are also useful in preparing PRO polypeptides, in assays to
CC   identify other proteins or molecules involved in binding reaction, to
CC   generate transgenic animals or knockout animals, which in turn are
CC   useful in the development and screening of therapeutically useful
CC   reagents, for chromosome identification, and tissue typing. The PRO
CC   polypeptides and nucleic acid molecules are also useful in gene
CC   therapy, and as molecular weight markers for protein electrophoresis
CC   purposes. Anti-PRO antibodies may be used in diagnostic assays for
CC   PRO polypeptides, or for the affinity purification of the polypeptides
CC   from recombinant cell culture or natural sources. The present sequence
CC   represents a cDNA encoding a human PRO polypeptide of the invention.
XX
SQ   Sequence 1977 BP; 513 A; 412 C; 482 G; 570 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 25;  Length 1977;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db       1639 GCAAACATCAAATAATATAA 1620


RESULT 38
ACA57934/c
ID   ACA57934 standard; cDNA; 1977 BP.
XX
AC   ACA57934;
XX
DT   10-JUN-2003  (first entry)
XX
DE   Human PRO4499 cDNA.
XX
KW   Human; PRO; secreted; transmembrane; cytostatic; TNF-alpha; blood; gene;
KW   tumour necrosis factor alpha release; chondrocyte cell; proliferation;
KW   differentiation; tumour; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   US2003036143-A1.
XX
PD   20-FEB-2003.
XX
PF   02-JUL-2002; 2002US-0187600.
XX
PR   16-SEP-1998;   98WO-US19330.
PR   07-OCT-1998;   98WO-US21141.
PR   01-DEC-1998;   98WO-US25108.
PR   08-MAR-1999;   99WO-US05028.
PR   14-MAY-1999;   99WO-US10733.
PR   02-JUN-1999;   99WO-US12252.
PR   01-SEP-1999;   99WO-US20111.
PR   15-SEP-1999;   99WO-US21090.
PR   01-DEC-1999;   99WO-US28301.
PR   02-DEC-1999;   99WO-US28551.
PR   30-DEC-1999;   99WO-US31274.
PR   05-JAN-2000; 2000WO-US00219.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US05004.
PR   01-MAR-2000; 2000WO-US05601.
PR   02-MAR-2000; 2000WO-US05841.
PR   15-MAR-2000; 2000WO-US06884.
PR   30-MAR-2000; 2000WO-US08439.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   28-JUL-2000; 2000WO-US20710.
PR   24-AUG-2000; 2000WO-US23328.
PR   08-NOV-2000; 2000WO-US30952.
PR   01-DEC-2000; 2000WO-US32678.
PR   20-DEC-2000; 2000WO-US34956.
PR   28-FEB-2001; 2001WO-US06520.
PR   01-JUN-2001; 2001WO-US17800.
PR   20-JUN-2001; 2001WO-US19692.
PR   29-JUN-2001; 2001WO-US21066.
PR   09-JUL-2001; 2001WO-US21735.
PR   29-AUG-2001; 2001WO-US27099.
PR   18-SEP-1997;   97US-059263P.
PR   18-SEP-1997;   97US-059266P.
PR   17-OCT-1997;   97US-062250P.
PR   21-OCT-1997;   97US-063486P.
PR   24-OCT-1997;   97US-063120P.
PR   24-OCT-1997;   97US-063121P.
PR   28-OCT-1997;   97US-063540P.
PR   28-OCT-1997;   97US-063541P.
PR   28-OCT-1997;   97US-063544P.
PR   28-OCT-1997;   97US-063564P.
PR   29-OCT-1997;   97US-063734P.
PR   31-OCT-1997;   97US-063870P.
PR   31-OCT-1997;   97US-064103P.
PR   13-NOV-1997;   97US-065311P.
PR   21-NOV-1997;   97US-066120P.
PR   24-NOV-1997;   97US-066466P.
PR   24-NOV-1997;   97US-066772P.
PR   11-DEC-1997;   97US-069335P.
PR   12-DEC-1997;   97US-069425P.
PR   17-DEC-1997;   97US-069870P.
PR   18-DEC-1997;   97US-068017P.
PR   10-MAR-1998;   98US-077450P.
PR   11-MAR-1998;   98US-077632P.
PR   11-MAR-1998;   98US-077649P.
PR   20-MAR-1998;   98US-078886P.
PR   20-MAR-1998;   98US-078939P.
PR   27-MAR-1998;   98US-079664P.
PR   27-MAR-1998;   98US-079786P.
PR   31-MAR-1998;   98US-080107P.
PR   31-MAR-1998;   98US-080194P.
PR   01-APR-1998;   98US-080327P.
PR   01-APR-1998;   98US-080333P.
PR   08-APR-1998;   98US-081049P.
PR   08-APR-1998;   98US-081070P.
PR   09-APR-1998;   98US-081195P.
PR   15-APR-1998;   98US-081838P.
PR   21-APR-1998;   98US-082568P.
PR   21-APR-1998;   98US-082569P.
PR   22-APR-1998;   98US-082704P.
PR   22-APR-1998;   98US-082797P.
PR   28-APR-1998;   98US-083322P.
PR   29-APR-1998;   98US-083495P.
PR   29-APR-1998;   98US-083496P.
PR   29-APR-1998;   98US-083499P.
PR   29-APR-1998;   98US-083559P.
PR   05-MAY-1998;   98US-084366P.
PR   06-MAY-1998;   98US-084414P.
PR   07-MAY-1998;   98US-084639P.
PR   07-MAY-1998;   98US-084640P.
PR   07-MAY-1998;   98US-084643P.
PR   15-MAY-1998;   98US-085579P.
PR   15-MAY-1998;   98US-085580P.
PR   15-MAY-1998;   98US-085582P.
PR   15-MAY-1998;   98US-085700P.
PR   18-MAY-1998;   98US-086023P.
PR   22-MAY-1998;   98US-086392P.
PR   22-MAY-1998;   98US-086486P.
PR   28-MAY-1998;   98US-087098P.
PR   28-MAY-1998;   98US-087208P.
```

PR 02-JUN-1998; 98US-087609P.
PR 02-JUN-1998; 98US-087759P.
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088029P.
PR 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088326P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088655P.
PR 10-JUN-1998; 98US-088722P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088740P.
PR 10-JUN-1998; 98US-088811P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088825P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088863P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089090P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089908P.
PR 19-JUN-1998; 98US-089952P.
PR 22-JUN-1998; 98US-090246P.
PR 22-JUN-1998; 98US-090252P.
PR 22-JUN-1998; 98US-090254P.
PR 24-JUN-1998; 98US-090429P.
PR 24-JUN-1998; 98US-090435P.
PR 24-JUN-1998; 98US-090444P.
PR 24-JUN-1998; 98US-090461P.
PR 24-JUN-1998; 98US-090535P.
PR 24-JUN-1998; 98US-090540P.
PR 25-JUN-1998; 98US-090676P.
PR 25-JUN-1998; 98US-090678P.
PR 25-JUN-1998; 98US-090688P.
PR 25-JUN-1998; 98US-090690P.
PR 25-JUN-1998; 98US-090694P.
PR 25-JUN-1998; 98US-090695P.
PR 25-JUN-1998; 98US-090696P.
PR 26-JUN-1998; 98US-090862P.
PR 26-JUN-1998; 98US-090863P.
PR 26-JUN-1998; 98US-091010P.
PR 01-JUL-1998; 98US-091359P.
PR 01-JUL-1998; 98US-091544P.
PR 02-JUL-1998; 98US-091478P.
PR 02-JUL-1998; 98US-091486P.
PR 02-JUL-1998; 98US-091626P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091632P.
PR 24-JUL-1998; 98US-094006P.
PR 04-AUG-1998; 98US-095282P.
PR 10-AUG-1998; 98US-095998P.
PR 10-AUG-1998; 98US-096012P.
PR 17-AUG-1998; 98US-096757P.
PR 17-AUG-1998; 98US-096766P.
PR 17-AUG-1998; 98US-096867P.
PR 17-AUG-1998; 98US-096891P.
PR 17-AUG-1998; 98US-096897P.
PR 18-AUG-1998; 98US-096949P.
PR 18-AUG-1998; 98US-096959P.
PR 18-AUG-1998; 98US-097022P.
PR 26-AUG-1998; 98US-097952P.
PR 26-AUG-1998; 98US-097954P.
PR 26-AUG-1998; 98US-097955P.
PR 26-AUG-1998; 98US-097971P.
PR 26-AUG-1998; 98US-097974P.
PR 26-AUG-1998; 98US-098014P.
PR 01-SEP-1998; 98US-098716P.
PR 01-SEP-1998; 98US-098723P.
PR 02-SEP-1998; 98US-098803P.
PR 02-SEP-1998; 98US-098821P.
PR 02-SEP-1998; 98US-098843P.
PR 09-SEP-1998; 98US-099602P.
PR 10-SEP-1998; 98US-099741P.
PR 10-SEP-1998; 98US-099754P.
PR 10-SEP-1998; 98US-099763P.
PR 10-SEP-1998; 98US-099812P.

Query Match 0.2%; Score 20; DB 25; Length 1977;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4861 GCAAACATCAAATAATATAA 4880
||||||||||||||||||||
Db 1639 GCAAACATCAAATAATATAA 1620

RESULT 39
ABX98404/c
ID ABX98404 standard; cDNA; 1977 BP.
XX
AC ABX98404;
XX
DT 19-MAY-2003 (first entry)
XX
DE Human cDNA encoding a secreted/transmembrane protein, SEQ ID 503.
XX
KW Human; ss; gene; PRO; secreted protein; transmembrane protein;
KW cytostatic; antiarthritic; osteopathic; adrenal tumour; lung tumour;
KW colon tumour; breast tumour; prostate tumour; rectal tumour;
KW cervical tumour; liver tumour; TNF-alpha release; arthritis;
KW tumour necrosis factor alpha; chondrocyte cell; bone disorder;
KW cartilage disorder; sports injury.
XX
OS Homo sapiens.
XX
PN US2003036156-A1.
XX
PD 20-FEB-2003.
XX
PF 02-JUL-2002; 2002US-0188767.
XX
PR 16-SEP-1998; 98WO-US19330.
PR 07-OCT-1998; 98WO-US21141.
PR 01-DEC-1998; 98WO-US25108.
PR 08-MAR-1999; 99WO-US05028.
PR 14-MAY-1999; 99WO-US10733.
PR 02-JUN-1999; 99WO-US12252.
PR 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 01-DEC-1999; 99WO-US28301.
PR 02-DEC-1999; 99WO-US28551.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US05004.
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 15-MAR-2000; 2000WO-US06884.
PR 30-MAR-2000; 2000WO-US08439.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.

PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 29-AUG-2001; 2001WO-US27099.
PR 18-SEP-1997; 97US-059263P.
PR 18-SEP-1997; 97US-059266P.
PR 17-OCT-1997; 97US-062250P.
PR 21-OCT-1997; 97US-063486P.
PR 24-OCT-1997; 97US-063120P.
PR 24-OCT-1997; 97US-063121P.
PR 28-OCT-1997; 97US-063540P.
PR 28-OCT-1997; 97US-063541P.
PR 28-OCT-1997; 97US-063544P.
PR 28-OCT-1997; 97US-063564P.
PR 29-OCT-1997; 97US-063734P.
PR 31-OCT-1997; 97US-063870P.
PR 31-OCT-1997; 97US-064103P.
PR 13-NOV-1997; 97US-065311P.
PR 21-NOV-1997; 97US-066120P.
PR 24-NOV-1997; 97US-066466P.
PR 24-NOV-1997; 97US-066772P.
PR 11-DEC-1997; 97US-069335P.
PR 12-DEC-1997; 97US-069425P.
PR 17-DEC-1997; 97US-069870P.
PR 18-DEC-1997; 97US-068017P.
PR 10-MAR-1998; 98US-077450P.
PR 11-MAR-1998; 98US-077632P.
PR 11-MAR-1998; 98US-077649P.
PR 20-MAR-1998; 98US-078886P.
PR 20-MAR-1998; 98US-078939P.
PR 27-MAR-1998; 98US-079664P.
PR 27-MAR-1998; 98US-079786P.
PR 31-MAR-1998; 98US-080107P.
PR 31-MAR-1998; 98US-080194P.
PR 01-APR-1998; 98US-080327P.
PR 01-APR-1998; 98US-080333P.
PR 08-APR-1998; 98US-081049P.
PR 08-APR-1998; 98US-081070P.
PR 09-APR-1998; 98US-081195P.
PR 15-APR-1998; 98US-081838P.
PR 21-APR-1998; 98US-082568P.
PR 21-APR-1998; 98US-082569P.
PR 22-APR-1998; 98US-082704P.
PR 22-APR-1998; 98US-082797P.
PR 28-APR-1998; 98US-083322P.
PR 29-APR-1998; 98US-083495P.
PR 29-APR-1998; 98US-083496P.
PR 29-APR-1998; 98US-083499P.
PR 29-APR-1998; 98US-083559P.
PR 05-MAY-1998; 98US-084366P.
PR 06-MAY-1998; 98US-084414P.
PR 07-MAY-1998; 98US-084639P.
PR 07-MAY-1998; 98US-084640P.
PR 07-MAY-1998; 98US-084643P.
PR 15-MAY-1998; 98US-085579P.
PR 15-MAY-1998; 98US-085580P.
PR 15-MAY-1998; 98US-085582P.
PR 15-MAY-1998; 98US-085700P.
PR 18-MAY-1998; 98US-086023P.
PR 22-MAY-1998; 98US-086392P.
PR 22-MAY-1998; 98US-086486P.
PR 28-MAY-1998; 98US-087098P.
PR 28-MAY-1998; 98US-087208P.
PR 02-JUN-1998; 98US-087609P.
PR 02-JUN-1998; 98US-087759P.
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088029P.
PR 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088326P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088655P.
PR 10-JUN-1998; 98US-088722P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088740P.
PR 10-JUN-1998; 98US-088811P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088825P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088863P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089090P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089908P.
PR 19-JUN-1998; 98US-089952P.
PR 22-JUN-1998; 98US-090246P.
PR 22-JUN-1998; 98US-090252P.
PR 22-JUN-1998; 98US-090254P.
PR 24-JUN-1998; 98US-090429P.
PR 24-JUN-1998; 98US-090435P.
PR 24-JUN-1998; 98US-090444P.
PR 24-JUN-1998; 98US-090461P.
PR 24-JUN-1998; 98US-090535P.
PR 24-JUN-1998; 98US-090540P.
PR 25-JUN-1998; 98US-090676P.
PR 25-JUN-1998; 98US-090678P.
PR 25-JUN-1998; 98US-090688P.
PR 25-JUN-1998; 98US-090690P.
PR 25-JUN-1998; 98US-090694P.
PR 25-JUN-1998; 98US-090695P.
PR 25-JUN-1998; 98US-090696P.
PR 26-JUN-1998; 98US-090862P.
PR 26-JUN-1998; 98US-090863P.
PR 26-JUN-1998; 98US-091010P.
PR 01-JUL-1998; 98US-091359P.
PR 01-JUL-1998; 98US-091544P.
PR 02-JUL-1998; 98US-091478P.
PR 02-JUL-1998; 98US-091486P.
PR 02-JUL-1998; 98US-091626P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091632P.
PR 24-JUL-1998; 98US-094006P.
PR 04-AUG-1998; 98US-095282P.
PR 10-AUG-1998; 98US-095998P.
PR 10-AUG-1998; 98US-096012P.
PR 17-AUG-1998; 98US-096757P.
PR 17-AUG-1998; 98US-096766P.
PR 17-AUG-1998; 98US-096867P.
PR 17-AUG-1998; 98US-096891P.
PR 17-AUG-1998; 98US-096897P.
PR 18-AUG-1998; 98US-096949P.
PR 18-AUG-1998; 98US-096959P.
PR 18-AUG-1998; 98US-097022P.
PR 26-AUG-1998; 98US-097952P.
PR 26-AUG-1998; 98US-097954P.
PR 26-AUG-1998; 98US-097955P.
PR 26-AUG-1998; 98US-097971P.
PR 26-AUG-1998; 98US-097974P.
PR 26-AUG-1998; 98US-098014P.
PR 01-SEP-1998; 98US-098716P.

```
PR   01-SEP-1998;   98US-098723P.
PR   02-SEP-1998;   98US-098803P.
PR   02-SEP-1998;   98US-098821P.
PR   02-SEP-1998;   98US-098843P.
PR   09-SEP-1998;   98US-099602P.
PR   10-SEP-1998;   98US-099741P.

  Query Match             0.2%;  Score 20;  DB 25;  Length 1977;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db       1639 GCAAACATCAAATAATATAA 1620


RESULT 40
ABX98906/c
ID   ABX98906 standard; cDNA; 1977 BP.
XX
AC   ABX98906;
XX
DT   20-MAY-2003  (first entry)
XX
DE   Novel human secreted and transmembrane protein PRO4499 cDNA.
XX
KW   Human; secreted protein; transmembrane protein; cytostatic;
KW   gene Therapy; TNF-Agonist-Alpha; chondrocyte stimulator; tumour;
KW   adrenal tumour; lung tumour; colon tumour; breast tumour;
KW   prostate tumour; rectal tumour; cervical tumour; liver tumour;
KW   gene; ss.
XX
OS   Homo sapiens.
XX
PN   US2003036157-A1.
XX
PD   20-FEB-2003.
XX
PF   02-JUL-2002; 2002US-0188769.
XX
PR   16-SEP-1998;   98WO-US19330.
PR   07-OCT-1998;   98WO-US21141.
PR   01-DEC-1998;   98WO-US25108.
PR   08-MAR-1999;   99WO-US05028.
PR   14-MAY-1999;   99WO-US10733.
PR   02-JUN-1999;   99WO-US12252.
PR   01-SEP-1999;   99WO-US20111.
PR   15-SEP-1999;   99WO-US21090.
PR   01-DEC-1999;   99WO-US28301.
PR   02-DEC-1999;   99WO-US28551.
PR   30-DEC-1999;   99WO-US31274.
PR   05-JAN-2000;  2000WO-US00219.
PR   18-FEB-2000;  2000WO-US04341.
PR   18-FEB-2000;  2000WO-US04342.
PR   22-FEB-2000;  2000WO-US04414.
PR   24-FEB-2000;  2000WO-US05004.
PR   01-MAR-2000;  2000WO-US05601.
PR   02-MAR-2000;  2000WO-US05841.
PR   15-MAR-2000;  2000WO-US06884.
PR   30-MAR-2000;  2000WO-US08439.
PR   17-MAY-2000;  2000WO-US13705.
PR   22-MAY-2000;  2000WO-US14042.
PR   30-MAY-2000;  2000WO-US14941.
PR   02-JUN-2000;  2000WO-US15264.
PR   28-JUL-2000;  2000WO-US20710.
PR   24-AUG-2000;  2000WO-US23328.
PR   08-NOV-2000;  2000WO-US30952.
PR   01-DEC-2000;  2000WO-US32678.
PR   20-DEC-2000;  2000WO-US34956.
PR   28-FEB-2001;  2001WO-US06520.
PR   01-JUN-2001;  2001WO-US17800.
PR   20-JUN-2001;  2001WO-US19692.
PR   29-JUN-2001;  2001WO-US21066.
PR   09-JUL-2001;  2001WO-US21735.
PR   29-AUG-2001;  2001WO-US27099.
PR   18-SEP-1997;   97US-059263P.
PR   18-SEP-1997;   97US-059266P.
PR   17-OCT-1997;   97US-062250P.
PR   21-OCT-1997;   97US-063486P.
PR   24-OCT-1997;   97US-063120P.
PR   24-OCT-1997;   97US-063121P.
PR   28-OCT-1997;   97US-063540P.
PR   28-OCT-1997;   97US-063541P.
PR   28-OCT-1997;   97US-063544P.
PR   28-OCT-1997;   97US-063564P.
PR   29-OCT-1997;   97US-063734P.
PR   31-OCT-1997;   97US-063870P.
PR   31-OCT-1997;   97US-064103P.
PR   13-NOV-1997;   97US-065311P.
PR   21-NOV-1997;   97US-066120P.
PR   24-NOV-1997;   97US-066466P.
PR   24-NOV-1997;   97US-066772P.
PR   11-DEC-1997;   97US-069335P.
PR   12-DEC-1997;   97US-069425P.
PR   17-DEC-1997;   97US-069870P.
PR   18-DEC-1997;   97US-068017P.
PR   10-MAR-1998;   98US-077450P.
PR   11-MAR-1998;   98US-077632P.
PR   11-MAR-1998;   98US-077649P.
PR   20-MAR-1998;   98US-078886P.
PR   20-MAR-1998;   98US-078939P.
PR   27-MAR-1998;   98US-079664P.
PR   27-MAR-1998;   98US-079786P.
PR   31-MAR-1998;   98US-080107P.
PR   31-MAR-1998;   98US-080194P.
PR   01-APR-1998;   98US-080327P.
PR   01-APR-1998;   98US-080333P.
PR   08-APR-1998;   98US-081049P.
PR   08-APR-1998;   98US-081070P.
PR   09-APR-1998;   98US-081195P.
PR   15-APR-1998;   98US-081838P.
PR   21-APR-1998;   98US-082568P.
PR   21-APR-1998;   98US-082569P.
PR   22-APR-1998;   98US-082704P.
PR   22-APR-1998;   98US-082797P.
PR   28-APR-1998;   98US-083322P.
PR   29-APR-1998;   98US-083495P.
PR   29-APR-1998;   98US-083496P.
PR   29-APR-1998;   98US-083499P.
PR   29-APR-1998;   98US-083559P.
PR   05-MAY-1998;   98US-084366P.
PR   06-MAY-1998;   98US-084414P.
PR   07-MAY-1998;   98US-084639P.
PR   07-MAY-1998;   98US-084640P.
PR   07-MAY-1998;   98US-084643P.
PR   15-MAY-1998;   98US-085579P.
PR   15-MAY-1998;   98US-085580P.
PR   15-MAY-1998;   98US-085582P.
PR   15-MAY-1998;   98US-085700P.
PR   18-MAY-1998;   98US-086023P.
PR   22-MAY-1998;   98US-086392P.
PR   22-MAY-1998;   98US-086486P.
PR   28-MAY-1998;   98US-087098P.
PR   28-MAY-1998;   98US-087208P.
PR   02-JUN-1998;   98US-087609P.
PR   02-JUN-1998;   98US-087759P.
PR   03-JUN-1998;   98US-087827P.
PR   04-JUN-1998;   98US-088025P.
PR   04-JUN-1998;   98US-088028P.
PR   04-JUN-1998;   98US-088029P.
PR   04-JUN-1998;   98US-088033P.
PR   04-JUN-1998;   98US-088326P.
PR   05-JUN-1998;   98US-088167P.
PR   05-JUN-1998;   98US-088202P.
PR   05-JUN-1998;   98US-088212P.
PR   05-JUN-1998;   98US-088217P.
```

```
PR   09-JUN-1998;   98US-088655P.
PR   10-JUN-1998;   98US-088722P.
PR   10-JUN-1998;   98US-088738P.
PR   10-JUN-1998;   98US-088740P.
PR   10-JUN-1998;   98US-088811P.
PR   10-JUN-1998;   98US-088824P.
PR   10-JUN-1998;   98US-088825P.
PR   10-JUN-1998;   98US-088826P.
PR   11-JUN-1998;   98US-088861P.
PR   11-JUN-1998;   98US-088863P.
PR   11-JUN-1998;   98US-088876P.
PR   12-JUN-1998;   98US-089090P.
PR   12-JUN-1998;   98US-089105P.
PR   16-JUN-1998;   98US-089512P.
PR   16-JUN-1998;   98US-089514P.
PR   17-JUN-1998;   98US-089538P.
PR   17-JUN-1998;   98US-089598P.
PR   17-JUN-1998;   98US-089653P.
PR   18-JUN-1998;   98US-089908P.
PR   19-JUN-1998;   98US-089952P.
PR   22-JUN-1998;   98US-090246P.
PR   22-JUN-1998;   98US-090252P.
PR   22-JUN-1998;   98US-090254P.
PR   24-JUN-1998;   98US-090429P.
PR   24-JUN-1998;   98US-090435P.
PR   24-JUN-1998;   98US-090444P.
PR   24-JUN-1998;   98US-090461P.
PR   24-JUN-1998;   98US-090535P.
PR   24-JUN-1998;   98US-090540P.
PR   25-JUN-1998;   98US-090676P.
PR   25-JUN-1998;   98US-090678P.
PR   25-JUN-1998;   98US-090688P.
PR   25-JUN-1998;   98US-090690P.
PR   25-JUN-1998;   98US-090694P.
PR   25-JUN-1998;   98US-090695P.
PR   25-JUN-1998;   98US-090696P.
PR   26-JUN-1998;   98US-090862P.
PR   26-JUN-1998;   98US-090863P.
PR   26-JUN-1998;   98US-091010P.
PR   01-JUL-1998;   98US-091359P.
PR   01-JUL-1998;   98US-091544P.
PR   02-JUL-1998;   98US-091478P.
PR   02-JUL-1998;   98US-091486P.
PR   02-JUL-1998;   98US-091626P.
PR   02-JUL-1998;   98US-091628P.
PR   02-JUL-1998;   98US-091632P.
PR   24-JUL-1998;   98US-094006P.
PR   04-AUG-1998;   98US-095282P.
PR   10-AUG-1998;   98US-095998P.
PR   10-AUG-1998;   98US-096012P.
PR   17-AUG-1998;   98US-096757P.
PR   17-AUG-1998;   98US-096766P.
PR   17-AUG-1998;   98US-096867P.
PR   17-AUG-1998;   98US-096891P.
PR   17-AUG-1998;   98US-096897P.
PR   18-AUG-1998;   98US-096949P.
PR   18-AUG-1998;   98US-096959P.
PR   18-AUG-1998;   98US-097022P.
PR   26-AUG-1998;   98US-097952P.
PR   26-AUG-1998;   98US-097954P.
PR   26-AUG-1998;   98US-097955P.
PR   26-AUG-1998;   98US-097971P.
PR   26-AUG-1998;   98US-097974P.
PR   26-AUG-1998;   98US-098014P.
PR   01-SEP-1998;   98US-098716P.
PR   01-SEP-1998;   98US-098723P.
PR   02-SEP-1998;   98US-098803P.
PR   02-SEP-1998;   98US-098821P.
PR   02-SEP-1998;   98US-098843P.
PR   09-SEP-1998;   98US-099602P.
PR   10-SEP-1998;   98US-099741P.
PR   10-SEP-1998;   98US-099754P.
```

```
  Query Match             0.2%;  Score 20;  DB 25;  Length 1977;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db       1639 GCAAACATCAAATAATATAA 1620


RESULT 41
ACA05951/c
ID   ACA05951 standard; cDNA; 1977 BP.
XX
AC   ACA05951;
XX
DT   29-MAY-2003  (first entry)
XX
DE   Human secreted/transmembrane protein (PRO) cDNA #252.
XX
KW   Human; gene; ss; secreted and transmembrane protein; PRO; TNF-alpha;
KW   tumour necrosis factor alpha; chondrocyte cell; tumour; gene therapy;
KW   tissue typing.
XX
OS   Homo sapiens.
XX
PN   US2003036162-A1.
XX
PD   20-FEB-2003.
XX
PF   12-JUL-2002; 2002US-0194423.
XX
PR   16-SEP-1998;   98WO-US19330.
PR   07-OCT-1998;   98WO-US21141.
PR   01-DEC-1998;   98WO-US25108.
PR   08-MAR-1999;   99WO-US05028.
PR   14-MAY-1999;   99WO-US10733.
PR   02-JUN-1999;   99WO-US12252.
PR   01-SEP-1999;   99WO-US20111.
PR   15-SEP-1999;   99WO-US21090.
PR   01-DEC-1999;   99WO-US28301.
PR   02-DEC-1999;   99WO-US28551.
PR   30-DEC-1999;   99WO-US31274.
PR   05-JAN-2000; 2000WO-US00219.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US05004.
PR   01-MAR-2000; 2000WO-US05601.
PR   02-MAR-2000; 2000WO-US05841.
PR   15-MAR-2000; 2000WO-US06884.
PR   30-MAR-2000; 2000WO-US08439.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   28-JUL-2000; 2000WO-US20710.
PR   24-AUG-2000; 2000WO-US23328.
PR   08-NOV-2000; 2000WO-US30952.
PR   01-DEC-2000; 2000WO-US32678.
PR   20-DEC-2000; 2000WO-US34956.
PR   28-FEB-2001; 2001WO-US06520.
PR   01-JUN-2001; 2001WO-US17800.
PR   20-JUN-2001; 2001WO-US19692.
PR   29-JUN-2001; 2001WO-US21066.
PR   09-JUL-2001; 2001WO-US21735.
PR   29-AUG-2001; 2001WO-US27099.
PR   26-JUN-1998;   98US-0105413.
PR   07-OCT-1998;   98US-0168978.
PR   06-NOV-1998;   98US-0187368.
PR   07-DEC-1998;   98US-0202054.
PR   03-MAR-1999;   99US-0254311.
PR   14-MAY-1999;   99US-0311832.
PR   14-MAY-1999;   99US-0380137.
```

```
PR  25-AUG-1999;   99US-0380138.
PR  25-AUG-1999;   99US-0380139.
PR  25-AUG-1999;   99US-0380142.
PR  18-OCT-1999;   99US-0403297.
PR  12-NOV-1999;   99US-0423844.
PR  22-AUG-2000; 2000US-0644848.
PR  18-SEP-2000; 2000US-0664610.
PR  18-SEP-2000; 2000US-0665350.
PR  08-NOV-2000; 2000US-0709238.
PR  20-DEC-2000; 2000US-0747259.
PR  22-MAR-2001; 2001US-0816744.
PR  10-MAY-2001; 2001US-0854208.
PR  10-MAY-2001; 2001US-0854280.
PR  25-MAY-2001; 2001US-0866028.
PR  05-JUN-2001; 2001US-0874503.
PR  18-JUL-2001; 2001US-0908827.
PR  30-JUL-2001; 2001US-0918585.
PR  06-AUG-2001; 2001US-0924419.
PR  13-AUG-2001; 2001US-0929404.
PR  16-AUG-2001; 2001US-0931836.
PR  28-AUG-2001; 2001US-0941992.
PR  04-SEP-2001; 2001US-0946374.
PR  15-JAN-2002; 2002US-0052586.
XX
PA  (GETH ) GENENTECH INC.
XX
PI  Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
PI  Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
DR  WPI; 2003-332039/31.
DR  P-PSDB; ABU67657.
XX
PT  New secreted and transmembrane PRO polypeptides and nucleic acids,
PT  useful in gene therapy, in chromosome and gene mapping, as chromosome
PT  markers, in tissue typing, and in chromosome identification
XX
PS  Claim 2; Fig 503; 706pp; English.
XX
CC  The invention discloses human nucleic acids encoding secreted and
CC  transmembrane (PRO) polypeptides. Also disclosed is an antibody that
CC  specifically binds to the PRO polypeptide, a method for stimulating the
CC  release of tumour necrosis factor alpha (TNF-alpha) from human blood by
CC  contacting the blood a PRO polypeptide, a method for stimulating the
CC  proliferation or differentiation of chondrocyte cells by contacting the
CC  cells with a PRO polypeptide, a method for detecting the presence of a
CC  tumour in a mammal and an oligonucleotide probe derived from any of the
CC  PRO nucleotide sequences. The nucleotide sequences are useful as probes,
CC  in chromosome and gene mapping, in generating antisense RNA and DNA, in
CC  preparing PRO polypeptides by recombinant techniques and in gene therapy
CC  (e.g. for replacement of defective gene). The PRO polypeptides are useful
CC  as molecular weight markers for protein electrophoresis purposes, for
CC  chromosome identification, as chromosome markers, as therapeutic agents,
CC  for stimulating the release of TNF-alpha from human blood, for
CC  stimulating the proliferation or differentiation of chondrocytes and
CC  detecting the presence of a tumour. The PRO polypeptides and nucleic
CC  acids may also be used diagnostically for tissue typing. The sequences
CC  presented in ACA05700-ACA06004 are the cDNAs encoding the PRO
CC  polypeptides of the invention.
XX
SQ  Sequence 1977 BP; 513 A; 412 C; 482 G; 570 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 25;  Length 1977;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db       1639 GCAAACATCAAATAATATAA 1620


RESULT 42
ABX97995/c
ID  ABX97995 standard; cDNA; 1977 BP.
XX
AC  ABX97995;
XX
DT  16-MAY-2003  (first entry)
XX
DE  Human PRO polynucleotide #252.
XX
KW  Human; PRO; gene; ss; cytostatic; chromosome mapping; gene mapping;
KW  protein electrophoresis; tumour necrosis factor-alpha; TNF-alpha; blood;
KW  chondrocyte differentiation; chondrocyte proliferation; tumour.
XX
OS  Homo sapiens.
XX
PN  US2003032102-A1.
XX
PD  13-FEB-2003.
XX
PF  17-JUN-2002; 2002US-0173697.
XX
PR  16-SEP-1998;   98WO-US19330.
PR  07-OCT-1998;   98WO-US21141.
PR  01-DEC-1998;   98WO-US25108.
PR  08-MAR-1999;   99WO-US05028.
PR  14-MAY-1999;   99WO-US10733.
PR  02-JUN-1999;   99WO-US12252.
PR  01-SEP-1999;   99WO-US20111.
PR  15-SEP-1999;   99WO-US21090.
PR  01-DEC-1999;   99WO-US28301.
PR  02-DEC-1999;   99WO-US28551.
PR  30-DEC-1999;   99WO-US31274.
PR  05-JAN-2000; 2000WO-US00219.
PR  18-FEB-2000; 2000WO-US04341.
PR  18-FEB-2000; 2000WO-US04342.
PR  22-FEB-2000; 2000WO-US04414.
PR  24-FEB-2000; 2000WO-US05004.
PR  01-MAR-2000; 2000WO-US05601.
PR  02-MAR-2000; 2000WO-US05841.
PR  15-MAR-2000; 2000WO-US06884.
PR  30-MAR-2000; 2000WO-US08439.
PR  17-MAY-2000; 2000WO-US13705.
PR  22-MAY-2000; 2000WO-US14042.
PR  30-MAY-2000; 2000WO-US14941.
PR  02-JUN-2000; 2000WO-US15264.
PR  28-JUL-2000; 2000WO-US20710.
PR  24-AUG-2000; 2000WO-US23328.
PR  08-NOV-2000; 2000WO-US30952.
PR  01-DEC-2000; 2000WO-US32678.
PR  20-DEC-2000; 2000WO-US34956.
PR  28-FEB-2001; 2001WO-US06520.
PR  01-JUN-2001; 2001WO-US17800.
PR  20-JUN-2001; 2001WO-US19692.
PR  29-JUN-2001; 2001WO-US21066.
PR  09-JUL-2001; 2001WO-US21735.
PR  29-AUG-2001; 2001WO-US27099.
PR  18-SEP-1997;   97US-059263P.
PR  18-SEP-1997;   97US-059266P.
PR  17-OCT-1997;   97US-062250P.
PR  21-OCT-1997;   97US-063486P.
PR  24-OCT-1997;   97US-063120P.
PR  24-OCT-1997;   97US-063121P.
PR  28-OCT-1997;   97US-063540P.
PR  28-OCT-1997;   97US-063541P.
PR  28-OCT-1997;   97US-063544P.
PR  28-OCT-1997;   97US-063564P.
PR  29-OCT-1997;   97US-063734P.
PR  31-OCT-1997;   97US-063870P.
PR  31-OCT-1997;   97US-064103P.
PR  13-NOV-1997;   97US-065311P.
PR  21-NOV-1997;   97US-066120P.
PR  24-NOV-1997;   97US-066466P.
PR  24-NOV-1997;   97US-066772P.
PR  11-DEC-1997;   97US-069335P.
PR  12-DEC-1997;   97US-069425P.
```

PR 17-DEC-1997; 97US-069870P.
PR 18-DEC-1997; 97US-068017P.
PR 10-MAR-1998; 98US-077450P.
PR 11-MAR-1998; 98US-077632P.
PR 11-MAR-1998; 98US-077649P.
PR 20-MAR-1998; 98US-078886P.
PR 20-MAR-1998; 98US-078939P.
PR 27-MAR-1998; 98US-079664P.
PR 27-MAR-1998; 98US-079786P.
PR 31-MAR-1998; 98US-080107P.
PR 31-MAR-1998; 98US-080194P.
PR 01-APR-1998; 98US-080327P.
PR 01-APR-1998; 98US-080333P.
PR 08-APR-1998; 98US-081049P.
PR 08-APR-1998; 98US-081070P.
PR 09-APR-1998; 98US-081195P.
PR 15-APR-1998; 98US-081838P.
PR 21-APR-1998; 98US-082568P.
PR 21-APR-1998; 98US-082569P.
PR 22-APR-1998; 98US-082704P.
PR 22-APR-1998; 98US-082797P.
PR 28-APR-1998; 98US-083322P.
PR 29-APR-1998; 98US-083495P.
PR 29-APR-1998; 98US-083496P.
PR 29-APR-1998; 98US-083499P.
PR 29-APR-1998; 98US-083559P.
PR 05-MAY-1998; 98US-084366P.
PR 06-MAY-1998; 98US-084414P.
PR 07-MAY-1998; 98US-084639P.
PR 07-MAY-1998; 98US-084640P.
PR 07-MAY-1998; 98US-084643P.
PR 15-MAY-1998; 98US-085579P.
PR 15-MAY-1998; 98US-085580P.
PR 15-MAY-1998; 98US-085582P.
PR 15-MAY-1998; 98US-085700P.
PR 18-MAY-1998; 98US-086023P.
PR 22-MAY-1998; 98US-086392P.
PR 22-MAY-1998; 98US-086486P.
PR 28-MAY-1998; 98US-087098P.
PR 28-MAY-1998; 98US-087208P.
PR 02-JUN-1998; 98US-087609P.
PR 02-JUN-1998; 98US-087759P.
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088029P.
PR 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088326P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088655P.
PR 10-JUN-1998; 98US-088722P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088740P.
PR 10-JUN-1998; 98US-088811P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088825P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088863P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089090P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089908P.
PR 19-JUN-1998; 98US-089952P.
PR 22-JUN-1998; 98US-090246P.
PR 22-JUN-1998; 98US-090252P.
PR 22-JUN-1998; 98US-090254P.
PR 24-JUN-1998; 98US-090429P.
PR 24-JUN-1998; 98US-090435P.
PR 24-JUN-1998; 98US-090444P.
PR 24-JUN-1998; 98US-090461P.
PR 24-JUN-1998; 98US-090535P.
PR 24-JUN-1998; 98US-090540P.
PR 25-JUN-1998; 98US-090676P.
PR 25-JUN-1998; 98US-090678P.
PR 25-JUN-1998; 98US-090688P.
PR 25-JUN-1998; 98US-090690P.
PR 25-JUN-1998; 98US-090694P.
PR 25-JUN-1998; 98US-090695P.
PR 25-JUN-1998; 98US-090696P.
PR 26-JUN-1998; 98US-090862P.
PR 26-JUN-1998; 98US-090863P.
PR 26-JUN-1998; 98US-091010P.
PR 01-JUL-1998; 98US-091359P.
PR 01-JUL-1998; 98US-091544P.
PR 02-JUL-1998; 98US-091478P.
PR 02-JUL-1998; 98US-091486P.
PR 02-JUL-1998; 98US-091626P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091632P.
PR 24-JUL-1998; 98US-094006P.
PR 04-AUG-1998; 98US-095282P.
PR 10-AUG-1998; 98US-095998P.
PR 10-AUG-1998; 98US-096012P.
PR 17-AUG-1998; 98US-096757P.
PR 17-AUG-1998; 98US-096766P.
PR 17-AUG-1998; 98US-096867P.
PR 17-AUG-1998; 98US-096891P.
PR 17-AUG-1998; 98US-096897P.
PR 18-AUG-1998; 98US-096949P.
PR 18-AUG-1998; 98US-096959P.
PR 18-AUG-1998; 98US-097022P.
PR 26-AUG-1998; 98US-097952P.
PR 26-AUG-1998; 98US-097954P.
PR 26-AUG-1998; 98US-097955P.
PR 26-AUG-1998; 98US-097971P.
PR 26-AUG-1998; 98US-097974P.
PR 26-AUG-1998; 98US-098014P.
PR 01-SEP-1998; 98US-098716P.
PR 01-SEP-1998; 98US-098723P.
PR 02-SEP-1998; 98US-098803P.
PR 02-SEP-1998; 98US-098821P.
PR 02-SEP-1998; 98US-098843P.
PR 09-SEP-1998; 98US-099602P.
PR 10-SEP-1998; 98US-099741P.
PR 10-SEP-1998; 98US-099754P.
PR 10-SEP-1998; 98US-099763P.
PR 10-SEP-1998; 98US-099812P.

Query Match 0.2%; Score 20; DB 25; Length 1977;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4861 GCAAACATCAAATAATATAA 4880
||||||||||||||||||||
Db 1639 GCAAACATCAAATAATATAA 1620

RESULT 43
ABX78779/c
ID ABX78779 standard; cDNA; 1977 BP.
XX
AC ABX78779;
XX
DT 15-APR-2003 (first entry)
XX
DE Human PRO polynucleotide #252.
XX

KW Human; PBO; gene; ss; cytostatic; tumour; cancer; breast; lung; stomach;
KW liver; dog; cat; cow; horse; sheep; pig; goat; rabbit; ADEPT;
KW antibody-dependent enzyme mediated prodrug therapy.
XX
OS Homo sapiens.
XX
PN US2003027272-A1.
XX
PD 06-FEB-2003.
XX
PF 21-JUN-2002; 2002US-0176492.
XX
PR 16-SEP-1998; 98WO-US19330.
PR 07-OCT-1998; 98WO-US21141.
PR 01-DEC-1998; 98WO-US25108.
PR 08-MAR-1999; 99WO-US05028.
PR 10-MAR-1999; 99WO-US05190.
PR 14-MAY-1999; 99WO-US10733.
PR 02-JUN-1999; 99WO-US12252.
PR 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28301.
PR 02-DEC-1999; 99WO-US28551.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US05004.
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 10-MAR-2000; 2000WO-US06319.
PR 15-MAR-2000; 2000WO-US06884.
PR 30-MAR-2000; 2000WO-US08439.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 10-NOV-2000; 2000WO-US30873.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 18-SEP-1997; 97US-059263P.
PR 18-SEP-1997; 97US-059266P.
PR 17-OCT-1997; 97US-062250P.
PR 21-OCT-1997; 97US-063486P.
PR 24-OCT-1997; 97US-063120P.
PR 24-OCT-1997; 97US-063121P.
PR 28-OCT-1997; 97US-063540P.
PR 28-OCT-1997; 97US-063541P.
PR 28-OCT-1997; 97US-063544P.
PR 28-OCT-1997; 97US-063564P.
PR 29-OCT-1997; 97US-063734P.
PR 31-OCT-1997; 97US-063870P.
PR 31-OCT-1997; 97US-064103P.
PR 13-NOV-1997; 97US-065311P.
PR 21-NOV-1997; 97US-066120P.
PR 24-NOV-1997; 97US-066466P.
PR 24-NOV-1997; 97US-066772P.
PR 11-DEC-1997; 97US-069335P.
PR 12-DEC-1997; 97US-069425P.
PR 17-DEC-1997; 97US-069870P.
PR 18-DEC-1997; 97US-068017P.
PR 10-MAR-1998; 98US-077450P.
PR 11-MAR-1998; 98US-077632P.
PR 11-MAR-1998; 98US-077649P.
PR 20-MAR-1998; 98US-078886P.
PR 20-MAR-1998; 98US-078939P.
PR 27-MAR-1998; 98US-079664P.
PR 27-MAR-1998; 98US-079786P.
PR 31-MAR-1998; 98US-080107P.
PR 31-MAR-1998; 98US-080194P.
PR 01-APR-1998; 98US-080327P.
PR 01-APR-1998; 98US-080333P.
PR 08-APR-1998; 98US-081049P.
PR 08-APR-1998; 98US-081070P.
PR 09-APR-1998; 98US-081195P.
PR 15-APR-1998; 98US-081838P.
PR 21-APR-1998; 98US-082568P.
PR 21-APR-1998; 98US-082569P.
PR 22-APR-1998; 98US-082704P.
PR 22-APR-1998; 98US-082797P.
PR 28-APR-1998; 98US-083322P.
PR 29-APR-1998; 98US-083495P.
PR 29-APR-1998; 98US-083496P.
PR 29-APR-1998; 98US-083499P.
PR 29-APR-1998; 98US-083559P.
PR 05-MAY-1998; 98US-084366P.
PR 06-MAY-1998; 98US-084414P.
PR 07-MAY-1998; 98US-084639P.
PR 07-MAY-1998; 98US-084640P.
PR 07-MAY-1998; 98US-084643P.
PR 15-MAY-1998; 98US-085579P.
PR 15-MAY-1998; 98US-085580P.
PR 15-MAY-1998; 98US-085582P.
PR 15-MAY-1998; 98US-085700P.
PR 18-MAY-1998; 98US-086023P.
PR 22-MAY-1998; 98US-086392P.
PR 22-MAY-1998; 98US-086486P.
PR 28-MAY-1998; 98US-087098P.
PR 28-MAY-1998; 98US-087208P.
PR 02-JUN-1998; 98US-087609P.
PR 02-JUN-1998; 98US-087759P.
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088029P.
PR 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088326P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088655P.
PR 10-JUN-1998; 98US-088722P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088740P.
PR 10-JUN-1998; 98US-088811P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088825P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088863P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089090P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089908P.
PR 19-JUN-1998; 98US-089952P.
PR 22-JUN-1998; 98US-090246P.
PR 22-JUN-1998; 98US-090252P.
PR 22-JUN-1998; 98US-090254P.
PR 24-JUN-1998; 98US-090429P.
PR 24-JUN-1998; 98US-090435P.

```
PR   24-JUN-1998;   98US-090444P.
PR   24-JUN-1998;   98US-090461P.
PR   24-JUN-1998;   98US-090535P.
PR   24-JUN-1998;   98US-090540P.
PR   25-JUN-1998;   98US-090676P.
PR   25-JUN-1998;   98US-090678P.
PR   25-JUN-1998;   98US-090688P.
PR   25-JUN-1998;   98US-090690P.
PR   25-JUN-1998;   98US-090694P.
PR   25-JUN-1998;   98US-090695P.
PR   25-JUN-1998;   98US-090696P.
PR   26-JUN-1998;   98US-090862P.
PR   26-JUN-1998;   98US-090863P.
PR   26-JUN-1998;   98US-091010P.
PR   01-JUL-1998;   98US-091359P.
PR   01-JUL-1998;   98US-091544P.
PR   02-JUL-1998;   98US-091478P.
PR   02-JUL-1998;   98US-091486P.
PR   02-JUL-1998;   98US-091626P.
PR   02-JUL-1998;   98US-091628P.
PR   02-JUL-1998;   98US-091632P.
PR   24-JUL-1998;   98US-094006P.
PR   04-AUG-1998;   98US-095282P.
PR   10-AUG-1998;   98US-095998P.
PR   10-AUG-1998;   98US-096012P.
PR   17-AUG-1998;   98US-096757P.
PR   17-AUG-1998;   98US-096766P.
PR   17-AUG-1998;   98US-096867P.
PR   17-AUG-1998;   98US-096891P.
PR   17-AUG-1998;   98US-096897P.
PR   18-AUG-1998;   98US-096949P.
PR   18-AUG-1998;   98US-096959P.
PR   18-AUG-1998;   98US-097022P.
PR   26-AUG-1998;   98US-097952P.
PR   26-AUG-1998;   98US-097954P.
PR   26-AUG-1998;   98US-097955P.
PR   26-AUG-1998;   98US-097971P.
PR   26-AUG-1998;   98US-097974P.
PR   26-AUG-1998;   98US-098014P.
PR   01-SEP-1998;   98US-098716P.
PR   01-SEP-1998;   98US-098723P.
PR   02-SEP-1998;   98US-098803P.
PR   02-SEP-1998;   98US-098821P.
PR   02-SEP-1998;   98US-098843P.
PR   09-SEP-1998;   98US-099602P.
PR   10-SEP-1998;   98US-099741P.

  Query Match             0.2%;  Score 20;  DB 25;  Length 1977;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db       1639 GCAAACATCAAATAATATAA 1620


RESULT 44
ABX75792/c
ID   ABX75792 standard; cDNA; 1977 BP.
XX
AC   ABX75792;
XX
DT   26-MAR-2003  (first entry)
XX
DE   Human cDNA encoding secreted/transmembrane protein, PRO4499.
XX
KW   Human; ss; gene; secreted protein; transmembrane protein; PRO;
KW   antiarthritic; vulnerary; tumour necrosis factor-alpha;
KW   chondrocyte cell proliferation; chondrocyte cell differentiation;
KW   tumour; adrenal tumour; lung tumour; colon tumour; breast tumour;
KW   prostate tumour; rectal tumour; cervical tumour; liver tumour;
KW   bone disorder; cartilage disorder; arthritis; sports injury.
XX
OS   Homo sapiens.
XX
PN   US2003022298-A1.
XX
PD   30-JAN-2003.
XX
PF   20-JUN-2002; 2002US-0176913.
XX
PR   05-NOV-1997;   97WO-US20069.
PR   10-SEP-1998;   98WO-US18824.
PR   14-SEP-1998;   98WO-US19177.
PR   16-SEP-1998;   98WO-US19330.
PR   17-SEP-1998;   98WO-US19437.
PR   07-OCT-1998;   98WO-US21141.
PR   20-NOV-1998;   98WO-US24855.
PR   01-DEC-1998;   98WO-US25108.
PR   05-JAN-1999;   99WO-US00106.
PR   08-MAR-1999;   99WO-US05028.
PR   10-MAR-1999;   99WO-US05190.
PR   20-APR-1999;   99WO-US08615.
PR   14-MAY-1999;   99WO-US10733.
PR   02-JUN-1999;   99WO-US12252.
PR   01-SEP-1999;   99WO-US20111.
PR   08-SEP-1999;   99WO-US20594.
PR   13-SEP-1999;   99WO-US20944.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   05-OCT-1999;   99WO-US23089.
PR   29-NOV-1999;   99WO-US28214.
PR   30-NOV-1999;   99WO-US28313.
PR   30-NOV-1999;   99WO-US28409.
PR   01-DEC-1999;   99WO-US28301.
PR   01-DEC-1999;   99WO-US28634.
PR   02-DEC-1999;   99WO-US28551.
PR   02-DEC-1999;   99WO-US28564.
PR   02-DEC-1999;   99WO-US28565.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   20-DEC-1999;   99WO-US30999.
PR   22-DEC-1999;   99WO-US30720.
PR   30-DEC-1999;   99WO-US31243.
PR   30-DEC-1999;   99WO-US31274.
PR   05-JAN-2000;  2000WO-US00219.
PR   06-JAN-2000;  2000WO-US00277.
PR   06-JAN-2000;  2000WO-US00376.
PR   11-FEB-2000;  2000WO-US03565.
PR   18-FEB-2000;  2000WO-US04341.
PR   18-FEB-2000;  2000WO-US04342.
PR   22-FEB-2000;  2000WO-US04414.
PR   24-FEB-2000;  2000WO-US05004.
PR   01-MAR-2000;  2000WO-US05601.
PR   02-MAR-2000;  2000WO-US05841.
PR   10-MAR-2000;  2000WO-US06319.
PR   15-MAR-2000;  2000WO-US06884.
PR   21-MAR-2000;  2000WO-US07532.
PR   30-MAR-2000;  2000WO-US08439.
PR   17-MAY-2000;  2000WO-US13705.
PR   22-MAY-2000;  2000WO-US14042.
PR   30-MAY-2000;  2000WO-US14941.
PR   02-JUN-2000;  2000WO-US15264.
PR   28-JUL-2000;  2000WO-US20710.
PR   11-AUG-2000;  2000WO-US22031.
PR   23-AUG-2000;  2000WO-US23522.
PR   24-AUG-2000;  2000WO-US23328.
PR   08-NOV-2000;  2000WO-US30952.
PR   10-NOV-2000;  2000WO-US30873.
PR   01-DEC-2000;  2000WO-US32678.
PR   20-DEC-2000;  2000WO-US34956.
PR   28-FEB-2001;  2001WO-US06520.
PR   22-MAR-2001;  2001WO-US09552.
PR   25-MAY-2001;  2001WO-US17092.
PR   01-JUN-2001;  2001WO-US17800.
PR   20-JUN-2001;  2001WO-US19692.
```

PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 29-AUG-2001; 2001WO-US27099.
PR 18-SEP-1997; 97US-059263P.
PR 18-SEP-1997; 97US-059266P.
PR 17-OCT-1997; 97US-062250P.
PR 21-OCT-1997; 97US-063486P.
PR 24-OCT-1997; 97US-063120P.
PR 24-OCT-1997; 97US-063121P.
PR 28-OCT-1997; 97US-063540P.
PR 28-OCT-1997; 97US-063541P.
PR 28-OCT-1997; 97US-063544P.
PR 28-OCT-1997; 97US-063564P.
PR 29-OCT-1997; 97US-063734P.
PR 31-OCT-1997; 97US-063870P.
PR 31-OCT-1997; 97US-064103P.
PR 13-NOV-1997; 97US-065311P.
PR 21-NOV-1997; 97US-066120P.
PR 24-NOV-1997; 97US-066466P.
PR 24-NOV-1997; 97US-066772P.
PR 11-DEC-1997; 97US-069335P.
PR 12-DEC-1997; 97US-069425P.
PR 17-DEC-1997; 97US-069870P.
PR 18-DEC-1997; 97US-068017P.
PR 10-MAR-1998; 98US-077450P.
PR 11-MAR-1998; 98US-077632P.
PR 11-MAR-1998; 98US-077649P.
PR 20-MAR-1998; 98US-078886P.
PR 20-MAR-1998; 98US-078939P.
PR 27-MAR-1998; 98US-079664P.
PR 27-MAR-1998; 98US-079786P.
PR 31-MAR-1998; 98US-080107P.
PR 31-MAR-1998; 98US-080194P.
PR 01-APR-1998; 98US-080327P.
PR 01-APR-1998; 98US-080333P.
PR 08-APR-1998; 98US-081049P.
PR 08-APR-1998; 98US-081070P.
PR 09-APR-1998; 98US-081195P.
PR 15-APR-1998; 98US-081838P.
PR 21-APR-1998; 98US-082568P.
PR 21-APR-1998; 98US-082569P.
PR 22-APR-1998; 98US-082704P.
PR 22-APR-1998; 98US-082797P.
PR 28-APR-1998; 98US-083322P.
PR 29-APR-1998; 98US-083495P.
PR 29-APR-1998; 98US-083496P.
PR 29-APR-1998; 98US-083499P.
PR 29-APR-1998; 98US-083559P.
PR 05-MAY-1998; 98US-084366P.
PR 06-MAY-1998; 98US-084414P.
PR 07-MAY-1998; 98US-084639P.
PR 07-MAY-1998; 98US-084640P.
PR 07-MAY-1998; 98US-084643P.
PR 15-MAY-1998; 98US-085579P.
PR 15-MAY-1998; 98US-085580P.
PR 15-MAY-1998; 98US-085582P.
PR 15-MAY-1998; 98US-085700P.
PR 18-MAY-1998; 98US-086023P.
PR 22-MAY-1998; 98US-086392P.
PR 22-MAY-1998; 98US-086486P.
PR 28-MAY-1998; 98US-087098P.
PR 28-MAY-1998; 98US-087208P.
PR 02-JUN-1998; 98US-087609P.
PR 02-JUN-1998; 98US-087759P.
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088029P.
PR 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088326P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088655P.
PR 10-JUN-1998; 98US-088722P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088740P.
PR 10-JUN-1998; 98US-088811P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088825P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088863P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089090P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089908P.
PR 19-JUN-1998; 98US-089952P.
PR 22-JUN-1998; 98US-090246P.
PR 22-JUN-1998; 98US-090252P.
PR 22-JUN-1998; 98US-090254P.
PR 24-JUN-1998; 98US-090429P.
PR 24-JUN-1998; 98US-090435P.
PR 24-JUN-1998; 98US-090444P.
PR 24-JUN-1998; 98US-090461P.
PR 24-JUN-1998; 98US-090535P.
PR 24-JUN-1998; 98US-090540P.
PR 25-JUN-1998; 98US-090676P.
PR 25-JUN-1998; 98US-090678P.
PR 25-JUN-1998; 98US-090688P.
PR 25-JUN-1998; 98US-090690P.
PR 25-JUN-1998; 98US-090694P.
PR 25-JUN-1998; 98US-090695P.
PR 25-JUN-1998; 98US-090696P.
PR 26-JUN-1998; 98US-090862P.
PR 26-JUN-1998; 98US-090863P.

Query Match 0.2%; Score 20; DB 25; Length 1977;
Best Local Similarity 100.0%; Pred.. No. 89;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4861 GCAAACATCAAATAATATAA 4880
||||||||||||||||||||
Db 1639 GCAAACATCAAATAATATAA 1620

RESULT 45
ABX76997/c
ID ABX76997 standard; cDNA; 1977 BP.
XX
AC ABX76997;
XX
DT 04-APR-2003 (first entry)
XX
DE Human PRO polynucleotide #252.
XX
KW Human; PRO; gene; ss; tumour necrosis factor-alpha; blood; cancer;
KW chondrocyte cell; tumour; adrenal tumour; lung; colon; breast; prostate;
KW kidney; rectum; cervix; liver; bone disorder; cartilage disorder;
KW arthritis; sports injury; genetic disorder; antiarthritic; vulnerary.
XX
OS Homo sapiens.
XX
PN US2003027280-A1.
XX
PD 06-FEB-2003.
XX
PF 20-JUN-2002; 2002US-0176993.
XX
PR 16-SEP-1998; 98WO-US19330.

PR 07-OCT-1998; 98WO-US21141.
PR 01-DEC-1998; 98WO-US25108.
PR 08-MAR-1999; 99WO-US05028.
PR 14-MAY-1999; 99WO-US10733.
PR 02-JUN-1999; 99WO-US12252.
PR 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 01-DEC-1999; 99WO-US28301.
PR 02-DEC-1999; 99WO-US28551.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US05004.
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 15-MAR-2000; 2000WO-US06884.
PR 30-MAR-2000; 2000WO-US08439.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 29-AUG-2001; 2001WO-US27099.
PR 18-SEP-1997; 97US-059263P.
PR 18-SEP-1997; 97US-059266P.
PR 17-OCT-1997; 97US-062250P.
PR 21-OCT-1997; 97US-063486P.
PR 24-OCT-1997; 97US-063120P.
PR 24-OCT-1997; 97US-063121P.
PR 28-OCT-1997; 97US-063540P.
PR 28-OCT-1997; 97US-063541P.
PR 28-OCT-1997; 97US-063544P.
PR 28-OCT-1997; 97US-063564P.
PR 29-OCT-1997; 97US-063734P.
PR 31-OCT-1997; 97US-063870P.
PR 31-OCT-1997; 97US-064103P.
PR 13-NOV-1997; 97US-065311P.
PR 21-NOV-1997; 97US-066120P.
PR 24-NOV-1997; 97US-066466P.
PR 24-NOV-1997; 97US-066772P.
PR 11-DEC-1997; 97US-069335P.
PR 12-DEC-1997; 97US-069425P.
PR 17-DEC-1997; 97US-069870P.
PR 18-DEC-1997; 97US-068017P.
PR 10-MAR-1998; 98US-077450P.
PR 11-MAR-1998; 98US-077632P.
PR 11-MAR-1998; 98US-077649P.
PR 20-MAR-1998; 98US-078886P.
PR 20-MAR-1998; 98US-078939P.
PR 27-MAR-1998; 98US-079664P.
PR 27-MAR-1998; 98US-079786P.
PR 31-MAR-1998; 98US-080107P.
PR 31-MAR-1998; 98US-080194P.
PR 01-APR-1998; 98US-080327P.
PR 01-APR-1998; 98US-080333P.
PR 08-APR-1998; 98US-081049P.
PR 08-APR-1998; 98US-081070P.
PR 09-APR-1998; 98US-081195P.
PR 15-APR-1998; 98US-081838P.
PR 21-APR-1998; 98US-082568P.
PR 21-APR-1998; 98US-082569P.
PR 22-APR-1998; 98US-082704P.
PR 22-APR-1998; 98US-082797P.
PR 28-APR-1998; 98US-083322P.
PR 29-APR-1998; 98US-083495P.
PR 29-APR-1998; 98US-083496P.
PR 29-APR-1998; 98US-083499P.
PR 29-APR-1998; 98US-083559P.
PR 05-MAY-1998; 98US-084366P.
PR 06-MAY-1998; 98US-084414P.
PR 07-MAY-1998; 98US-084639P.
PR 07-MAY-1998; 98US-084640P.
PR 07-MAY-1998; 98US-084643P.
PR 15-MAY-1998; 98US-085579P.
PR 15-MAY-1998; 98US-085580P.
PR 15-MAY-1998; 98US-085582P.
PR 15-MAY-1998; 98US-085700P.
PR 18-MAY-1998; 98US-086023P.
PR 22-MAY-1998; 98US-086392P.
PR 22-MAY-1998; 98US-086486P.
PR 28-MAY-1998; 98US-087098P.
PR 28-MAY-1998; 98US-087208P.
PR 02-JUN-1998; 98US-087609P.
PR 02-JUN-1998; 98US-087759P.
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088029P.
PR 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088326P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088655P.
PR 10-JUN-1998; 98US-088722P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088740P.
PR 10-JUN-1998; 98US-088811P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088825P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088863P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089090P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089908P.
PR 19-JUN-1998; 98US-089952P.
PR 22-JUN-1998; 98US-090246P.
PR 22-JUN-1998; 98US-090252P.
PR 22-JUN-1998; 98US-090254P.
PR 24-JUN-1998; 98US-090429P.
PR 24-JUN-1998; 98US-090435P.
PR 24-JUN-1998; 98US-090444P.
PR 24-JUN-1998; 98US-090461P.
PR 24-JUN-1998; 98US-090535P.
PR 24-JUN-1998; 98US-090540P.
PR 25-JUN-1998; 98US-090676P.
PR 25-JUN-1998; 98US-090678P.
PR 25-JUN-1998; 98US-090688P.
PR 25-JUN-1998; 98US-090690P.
PR 25-JUN-1998; 98US-090694P.
PR 25-JUN-1998; 98US-090695P.
PR 25-JUN-1998; 98US-090696P.
PR 26-JUN-1998; 98US-090862P.
PR 26-JUN-1998; 98US-090863P.
PR 26-JUN-1998; 98US-091010P.
PR 01-JUL-1998; 98US-091359P.
PR 01-JUL-1998; 98US-091544P.

PR 02-JUL-1998; 98US-091478P.
PR 02-JUL-1998; 98US-091486P.
PR 02-JUL-1998; 98US-091626P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091632P.
PR 24-JUL-1998; 98US-094006P.
PR 04-AUG-1998; 98US-095282P.
PR 10-AUG-1998; 98US-095998P.
PR 10-AUG-1998; 98US-096012P.
PR 17-AUG-1998; 98US-096757P.
PR 17-AUG-1998; 98US-096766P.
PR 17-AUG-1998; 98US-096867P.
PR 17-AUG-1998; 98US-096891P.
PR 17-AUG-1998; 98US-096897P.
PR 18-AUG-1998; 98US-096949P.
PR 18-AUG-1998; 98US-096959P.
PR 18-AUG-1998; 98US-097022P.
PR 26-AUG-1998; 98US-097952P.
PR 26-AUG-1998; 98US-097954P.
PR 26-AUG-1998; 98US-097955P.
PR 26-AUG-1998; 98US-097971P.
PR 26-AUG-1998; 98US-097974P.
PR 26-AUG-1998; 98US-098014P.
PR 01-SEP-1998; 98US-098716P.
PR 01-SEP-1998; 98US-098723P.
PR 02-SEP-1998; 98US-098803P.
PR 02-SEP-1998; 98US-098821P.
PR 02-SEP-1998; 98US-098843P.
PR 09-SEP-1998; 98US-099602P.
PR 10-SEP-1998; 98US-099741P.
PR 10-SEP-1998; 98US-099754P.
PR 10-SEP-1998; 98US-099763P.

Query Match 0.2%; Score 20; DB 25; Length 1977;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4861 GCAAACATCAAATAATATAA 4880
||||||||||||||||||||
Db 1639 GCAAACATCAAATAATATAA 1620

RESULT 46
ABX16837/c
ID ABX16837 standard; cDNA; 1977 BP.
XX
AC ABX16837;
XX
DT 03-FEB-2003 (first entry)
XX
DE Human cDNA encoding secreted/transmembrane protein #252.
XX
KW Human; ss; gene; secreted and transmembrane protein; blood;
KW tumour necrosis factor-alpha; chondrocyte cell proliferation;
KW chondrocyte cell differentiation; tumour; adrenal tumour; lung tumour;
KW colon tumour; breast tumour; prostate tumour; rectal tumour;
KW cervical tumour; liver tumour; bone disorder; cartilage disorder;
KW arthritis; sports injury.
XX
OS Homo sapiens.
XX
PN US2002127584-A1.
XX
PD 12-SEP-2002.
XX
PF 15-JAN-2002; 2002US-0052586.
XX
PR 16-SEP-1998; 98WO-US19330.
PR 07-OCT-1998; 98WO-US21141.
PR 01-DEC-1998; 98WO-US25108.
PR 06-JAN-1999; 2000WO-US00219.
PR 08-MAR-1999; 99WO-US05028.
PR 14-MAY-1999; 99WO-US10733.
PR 02-JUN-1999; 99WO-US12252.
PR 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 01-DEC-1999; 99WO-US28301.
PR 02-DEC-1999; 99WO-US28551.
PR 30-DEC-1999; 99WO-US31274.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US05004.
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 15-MAR-2000; 2000WO-US06884.
PR 30-MAR-2000; 2000WO-US08439.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 29-AUG-2001; 2001WO-US27099.
PR 18-SEP-1997; 97US-059263P.
PR 18-SEP-1997; 97US-059266P.
PR 17-OCT-1997; 97US-062250P.
PR 21-OCT-1997; 97US-063486P.
PR 24-OCT-1997; 97US-063120P.
PR 24-OCT-1997; 97US-063121P.
PR 28-OCT-1997; 97US-063540P.
PR 28-OCT-1997; 97US-063541P.
PR 28-OCT-1997; 97US-063544P.
PR 28-OCT-1997; 97US-063564P.
PR 29-OCT-1997; 97US-063734P.
PR 31-OCT-1997; 97US-063870P.
PR 31-OCT-1997; 97US-064103P.
PR 13-NOV-1997; 97US-065311P.
PR 21-NOV-1997; 97US-066120P.
PR 24-NOV-1997; 97US-066466P.
PR 24-NOV-1997; 97US-066772P.
PR 11-DEC-1997; 97US-069335P.
PR 12-DEC-1997; 97US-069425P.
PR 17-DEC-1997; 97US-069870P.
PR 18-DEC-1997; 97US-068017P.
PR 10-MAR-1998; 98US-077450P.
PR 11-MAR-1998; 98US-077632P.
PR 11-MAR-1998; 98US-077649P.
PR 20-MAR-1998; 98US-078886P.
PR 20-MAR-1998; 98US-078939P.
PR 27-MAR-1998; 98US-079664P.
PR 27-MAR-1998; 98US-079786P.
PR 31-MAR-1998; 98US-080107P.
PR 31-MAR-1998; 98US-080194P.
PR 01-APR-1998; 98US-080327P.
PR 01-APR-1998; 98US-080333P.
PR 08-APR-1998; 98US-081049P.
PR 08-APR-1998; 98US-081070P.
PR 09-APR-1998; 98US-081195P.
PR 15-APR-1998; 98US-081838P.
PR 21-APR-1998; 98US-082568P.
PR 21-APR-1998; 98US-082569P.
PR 22-APR-1998; 98US-082704P.
PR 22-APR-1998; 98US-082797P.
PR 28-APR-1998; 98US-083322P.
PR 29-APR-1998; 98US-083495P.
PR 29-APR-1998; 98US-083496P.
PR 29-APR-1998; 98US-083499P.

```
PR   29-APR-1998;   98US-083559P.
PR   05-MAY-1998;   98US-084366P.
PR   06-MAY-1998;   98US-084414P.
PR   07-MAY-1998;   98US-084639P.
PR   07-MAY-1998;   98US-084640P.
PR   07-MAY-1998;   98US-084643P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
PI   Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
DR   WPI; 2003-066893/06.
DR   P-PSDB; ABU10761.
XX
PT   Novel isolated PRO polypeptides e.g., PRO1079, PRO827, PRO791, PRO1131,
PT   PRO1316, PRO1183, PRO1343, PRO1760, PRO1567 or PRO4333, useful for
PT   stimulating release of tumor necrosis factor-alpha from human blood -
XX
PS   Claim 2; Fig 503; 701pp; English.
XX
CC   The invention relates to an isolated PRO polypeptide comprising at least
CC   80% sequence identity to the protein sequences appearing as ABU10510-
CC   ABU10814 (including a version lacking its associated signal peptide, or
CC   an isolated extracellular domain of a PRO polypeptide with or without
CC   its associated signal peptide. Also included are the nucleic acids
CC   encoding the PRO proteins (being secreted and transmembrane proteins)
CC   appearing as ABX16586-ABX16590, PRO expression vectors, host cells,
CC   chimaeric PRO fusion proteins, an anti-PRO antibody and a PRO
CC   derived oligonucleotide sequence. The PRO polypeptides are useful for
CC   stimulating release of tumour necrosis factor-alpha from human blood.
CC   The PRO polypeptide PRO6029 is useful for stimulating proliferation or
CC   differentiation of chondrocyte cells. The PRO polypeptides as specified
CC   in the specification and having differential expression in tumour cells,
CC   are useful for detecting presence of tumour in a mammal (such as adrenal
CC   tumour, lung tumour, colon tumour, breast tumour, prostate tumour, rectal
CC   tumour, cervical tumour or liver tumour. The PRO polypeptide PRO6029 is
CC   useful for treating various bone and/or cartilage disorders such as
CC   arthritis, and sports injuries. The PRO polypeptides are useful for
CC   screening compounds to identify ant/agonists. PRO nucleic acids
CC   are useful as hybridisation probes, in chromosome and gene mapping,
CC   in the generation of anti-sense RNA and DNA, for the preparation of PRO
CC   polypeptides and for generating knock-out animals. The present
CC   sequence encodes a PRO polypeptide.
XX
SQ   Sequence 1977 BP; 513 A; 412 C; 482 G; 570 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 25;  Length 1977;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db       1639 GCAAACATCAAATAATATAA 1620


RESULT 47
AAZ51194/c
ID   AAZ51194 standard; cDNA; 1981 BP.
XX
AC   AAZ51194;
XX
DT   06-JUN-2000  (first entry)
XX
DE   Human ischaemic heart disease associated protein PGP980.2 encoding cDNA.
XX
KW   Ischaemia; heart disease; human; PGP980.2; cardiomyopathy; cardiant;
KW   modulator; treat; prevention; diagnosis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             133..1152
FT                   /*tag=  a
FT                   /product= "Human ischaemic heart disease associated
FT                   protein, PGP980.2"
FT   sig_peptide     133..219
FT                   /*tag=  b
FT   mat_peptide     220..1149
FT                   /*tag=  c
FT                   /product= "Mature human ischaemic heart disease
FT                   associated protein, PGP980.2"
XX
PN   WO200011942-A1.
XX
PD   09-MAR-2000.
XX
PF   01-SEP-1999;   99WO-US20015.
XX
PR   01-SEP-1998;   98US-0098683.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Einstein R;
XX
DR   WPI; 2000-237720/20.
DR   P-PSDB; AAY70118.
XX
PT   Isolated nucleic acids encoding a proteins that is differentially
PT   expressed during ischemic heart disease and ischemic cardiomyopathy,
PT   useful for preventing, diagnosing and treating cardiac disease -
XX
PS   Claim 2; Page 51-54; 61pp; English.
XX
CC   The present sequence is the human cDNA encoding a protein designated as
CC.  PGP980.2, that is differentially expressed in ischaemic heart disorders.
CC   PGP980.2 DNA is isolated from human heart cDNA library and is expressed
CC   in heart, brain, placenta, lung, skeletal muscle, kidney and pancreas.
CC   This sequence has cardiant activity. The DNA may be used as probes to
CC   detect and quantify homologous sequences in samples. PGP980.2 sequences
CC   can be used to identify modulators, that can be used for treatment,
CC   diagnosis and prevention of diseases associated with inappropriate
CC   expression and activity of PGP980.2, such as ischaemic heart disease and
CC   ischaemic cardiomyopathy.
XX
SQ   Sequence 1981 BP; 552 A; 402 C; 462 G; 565 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 21;  Length 1981;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db       1587 GCAAACATCAAATAATATAA 1568


RESULT 48
ABA09180
ID   ABA09180 standard; cDNA; 1999 BP.
XX
AC   ABA09180;
XX
DT   11-JAN-2002  (first entry)
XX
DE   Human secreted protein homologue-encoding cDNA, SEQ ID NO:956.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW   haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW   inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW   proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW   myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW   chronic inflammatory condition; proliferative retinopathy;
KW   atherosclerosis; coronary heart disease; arterial ischaemia;
KW   bone disorder; osteoporosis; vascular growth disorder;
KW   tissue regeneration; wound healing; infection; immune disorder;
KW   cell culture; drug screening; gene therapy; antiinflammatory;
```

```
KW  antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW  cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW  antifungal; vulnerary; antiulcer; ss.
XX
OS  Homo sapiens.
XX
PN  WO200157188-A2.
XX
PD  09-AUG-2001.
XX
PF  05-FEB-2001; 2001WO-US03800.
XX
PR  03-FEB-2000; 2000US-0496914.
PR  27-APR-2000; 2000US-0560875.
XX
PA  (HYSE-) HYSEQ INC.
XX
PI  Tang YT,  Liu C,  Drmanac RT;
XX
DR  WPI; 2001-457740/49.
DR  P-PSDB; ABB11936.
XX
PT  Human proteins and DNA encoding sequences useful for preventing,
PT  treating or ameliorating a medical condition in a mammalian subject
PT  e.g. arthritis and cancer -
XX
PS  Claim 1; Page 817; 1963pp; English.
XX
CC  Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC  sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC  invention also relates to vectors and recombinant host cells comprising a
CC  nucleotide of the invention, methods of producing the novel polypeptides,
CC  antibodies against the polypeptides, methods of detecting the nucleotides
CC  or polypeptides in a sample, and methods of identifying compounds which
CC  bind to polypeptides of the invention. Although novel, many of the
CC  polypeptides of the invention have homology to known proteins, thereby
CC  giving an insight into their probable biological activities, and hence
CC  potential therapeutic applications. The polypeptides of the invention may
CC  have various activities, including cytokine, cell proliferation or cell
CC  differentiation activities; stem cell growth factor activity;
CC  haematopoiesis regulatory activity; tissue growth activity;
CC  immunomodulatory activity; activin- or inhibin-related activities;
CC  chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC  thrombolytic activities; receptor or ligand activities; or may be
CC  involved in oncogenesis, cancer cell proliferation or metastasis.
CC  Depending on their biological activities, polypeptides and nucleotides of
CC  the invention are useful for preventing, treating or ameliorating medical
CC  conditions, e.g., by protein or gene therapy. Such conditions include
CC  cancers, haematopoietic disorders (e.g.,myeloid or lymphoid cell
CC  disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC  proliferative retinopathy, atherosclerosis, coronary heart disease,
CC  arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC  vascular growth. Polypeptides involved with tissue regeneration and
CC  repair (or nucleic acids encoding them) may be used to promote wound
CC  healing (e.g., of burns, incisions and ulcers), while those with
CC  immunomodulatory activities may be used in the treatment of viral,
CC  bacterial and fungal infections in addition to immune disorders.
CC  Polypeptides with growth factor activity may be used in cell cultures to
CC  promote cell growth. For example, such polypeptides may be used to
CC  manipulate stem cells in culture to give rise to neuroepithelial cells
CC  that can be used to augment or replace cells damaged by illness,
CC  autoimmune disease or accidental damage. The polypeptides and nucleotides
CC  may also be used in the diagnosis of the above conditions, and in drug
CC  screening techniques. The present sequence represents a cDNA encoding a
CC  novel human polypeptide of the invention.
XX
SQ  Sequence 1999 BP; 571 A; 482 C; 415 G; 531 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 22;  Length 1999;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db        360 GCAAACATCAAATAATATAA 379


RESULT 49
AAK52615
ID  AAK52615 standard; cDNA; 1999 BP.
XX
AC  AAK52615;
XX
DT  06-NOV-2001  (first entry)
XX
DE  Human polynucleotide SEQ ID NO 2144.
XX
KW  Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW  vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW  tissue growth factor; immunomodulatory; cancer; leukaemia;
KW  nervous system disorder; arthritis; inflammation; ss.
XX
OS  Homo sapiens.
XX
PN  WO200157190-A2.
XX
PD  09-AUG-2001.
XX
PF  05-FEB-2001; 2001WO-US04098.
XX
PR  03-FEB-2000; 2000US-0496914.
PR  27-APR-2000; 2000US-0560875.
PR  20-JUN-2000; 2000US-0598075.
PR  19-JUL-2000; 2000US-0620325.
PR  01-SEP-2000; 2000US-0654936.
PR  15-SEP-2000; 2000US-0663561.
PR  20-OCT-2000; 2000US-0693325.
PR  30-NOV-2000; 2000US-0728422.
XX
PA  (HYSE-) HYSEQ INC.
XX
PI  Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y,  Ma Y;
PI  Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI  Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR  WPI; 2001-476283/51.
DR  P-PSDB; AAM79482.
XX
PT  Nucleic acids encoding polypeptides with cytokine-like activities,
PT  useful in diagnosis and gene therapy -
XX
PS  Claim 1; Page 4514; 6221pp; English.
XX
CC  The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC  encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC  cytokine, cell proliferation or cell differentiation or which may induce
CC  production of other cytokines in other cell populations. The
CC  polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC  peptide therapy. The polypeptides have various cytokine-like activities,
CC  e.g. stem cell growth factor activity, haematopoiesis regulating
CC  activity, tissue growth factor activity, immunomodulatory activity and
CC  activin/inhibin activity and may be useful in the diagnosis and/or
CC  treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC  inflammation.
CC  Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC  (AAM80020) are omitted as the relevant pages from the sequence listing
CC  were missing at the time of publication.
XX
SQ  Sequence 1999 BP; 571 A; 482 C; 415 G; 531 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 22;  Length 1999;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
```

```
Db        360 GCAAACATCAAATAATATAA 379


RESULT 50
AAH99744/c
ID   AAH99744 standard; cDNA; 2003 BP.
XX
AC   AAH99744;
XX
DT   16-OCT-2001  (first entry)
XX
DE   Human protein encoding cDNA sequence SEQ ID NO:579.
XX
KW   Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW   antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW   antibacterial; endocrine; cardiant; central nervous system; virucide;
KW   anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW   antiaggregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;
KW   dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW   neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW   immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW   antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW   cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW   genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW   thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW   allergic rhinitis; diabetes; multiple sclerosis; depression;
KW   Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW   neurological disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153455-A2.
XX
PD   26-JUL-2001.
XX
PF   22-DEC-2000; 2000WO-US35017.
XX
PR   23-DEC-1999;   99US-0471275.
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-457603/49.
DR   P-PSDB; AAM25803.
XX
PT   Isolated human polynucleotides encoding polypeptides, useful for the
PT   treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
PS   Claim 1; Page 620; 1217pp; English.
XX
CC   AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC   AAM25963. The proteins can have activities based on the tissues and
CC   cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC   antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC   central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC   cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;
CC   antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC   antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC   antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC   encoding them can be used in gene therapy, antisense therapy and vaccine
CC   production, The proteins and polynucleotides are useful for screening for
CC   agonists or antagonists of a protein and for the treatment and diagnosis
CC   of disorders associated with the activity of a protein e.g. inflammation;
CC   rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC   neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC   infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC   anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC   osteoporosis, severe combined immunodeficiency, eczema, allergic
CC   rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC   Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC   neurological disorders.
XX
SQ   Sequence 2003 BP; 531 A; 415 C; 486 G; 571 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 22;  Length 2003;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db       1640 GCAAACATCAAATAATATAA 1621


RESULT 51
AAZ51193/c
ID   AAZ51193 standard; cDNA; 2008 BP.
XX
AC   AAZ51193;
XX
DT   06-JUN-2000  (first entry)
XX
DE   Human ischaemic heart disease associated protein PGP980.1 encoding cDNA.
XX
KW   Ischaemia; heart disease; human; PGP980.1; cardiomyopathy; cardiant;
KW   modulator; treat; prevention; diagnosis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             184..1203
FT                   /*tag= a
FT                   /product= "Human ischaemic heart disease associated
FT                   protein, PGP980.1"
FT   sig_peptide     184..270
FT                   /*tag= b
FT   mat_peptide     271..1200
FT                   /*tag= c
FT                   /product= "Mature human ischaemic heart disease
FT                   associated protein, PGP980.1"
XX
PN   WO200011942-A1.
XX
PD   09-MAR-2000.
XX
PF   01-SEP-1999;   99WO-US20015.
XX
PR   01-SEP-1998;   98US-0098683.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Einstein R;
XX
DR   WPI; 2000-237720/20.
DR   P-PSDB; AAY70117.
XX
PT   Isolated nucleic acids encoding a proteins that is differentially
PT   expressed during ischemic heart disease and ischemic cardiomyopathy,
PT   useful for preventing, diagnosing and treating cardiac disease -
XX
PS   Claim 2; Page 47-50; 61pp; English.
XX
CC   The present sequence is the human cDNA encoding a protein designated as
CC   PGP980.1, that is differentially expressed in ischaemic heart disorders.
CC   PGP980.1 DNA is isolated from human heart cDNA library and is expressed
CC   in heart, brain, placenta, lung, skeletal muscle, kidney and pancreas.
CC   This sequence has cardiant activity. The DNA may be used as probes to
CC   detect and quantify homologous sequences in samples. PGP980.1 sequences
CC   can be used to identify modulators, that can be used for treatment,
CC   diagnosis and prevention of diseases associated with inappropriate
CC   expression and activity of PGP980.1, such as ischaemic heart disease
CC   and ischaemic cardiomyopathy.
XX
SQ   Sequence 2008 BP; 538 A; 414 C; 485 G; 571 T; 0 other;
```

```
  Query Match             0.2%;  Score 20;  DB 21;  Length 2008;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db       1638 GCAAACATCAAATAATATAA 1619


RESULT 52
AAV08830/c
ID   AAV08830 standard; cDNA; 2017 BP.
XX
AC   AAV08830;
XX
DT   19-FEB-1999  (first entry)
XX
DE   Gene No. 20 encoding human secreted protein.
XX
KW   Secreted protein; human; protein therapy; gene therapy; blood disorder;
KW   pathological condition; diagnosis; cancer; neurological disorder;
KW   developmental abnormality; foetal deficiency; leukaemia; hepatic disease;
KW   immune system disorder; Alzheimer's disease; cognitive disorder;
KW   schizophrenia; prostate disease; autoimmune disorder; AIDS; ds.
XX
OS   Homo sapiens.
XX
PN   WO9854206-A1.
XX
PD   03-DEC-1998.
XX
PF   28-MAY-1998;   98WO-US10868.
XX
PR   29-AUG-1997;   97US-0056296.
PR   30-MAY-1997;   97US-0044039.
PR   30-MAY-1997;   97US-0048093.
PR   30-MAY-1997;   97US-0048101.
PR   30-MAY-1997;   97US-0048190.
PR   30-MAY-1997;   97US-0048356.
PR   30-MAY-1997;   97US-0050935.
PR   29-AUG-1997;   97US-0056250.
PR   29-AUG-1997;   97US-0056293.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Carter KC,  Dillon PJ,  Endress GA,  Feng P,  Ni J;
PI   Rosen CA,  Ruben SM,  Yu G;
XX
DR   WPI; 1999-070209/06.
DR   P-PSDB; AAW73416.
XX
PT   New isolated human genes - useful for diagnosis and treatment of,
PT   e.g. cancers, neurological disorders, immune diseases, developmental
PT   disorders or blood disorders
XX
PS   Claim 1; Page 129-130; 188pp; English.
XX
CC   This sequence represents a cDNA of the invention, and is designated
CC   Gene No. 20. This sequence encodes a human secreted protein, and is
CC   expressed ubiquitously, including T-cells and amygdala.
CC   The DNA sequences of the invention and their corresponding secreted
CC   polypeptides are useful for preventing, treating or ameliorating medical
CC   conditions, e.g. by protein or gene therapy. Also pathological conditions
CC   can be diagnosed by determining the amount of the new polypeptides in a
CC   sample or by determining the presence of mutations in the DNA sequences.
CC   Specific uses are described for each of the DNA sequences and the encoded
CC   proteins, based on which tissues they are most highly expressed in, and
CC   include developing products for the diagnosis or treatment of cancer,
CC   tumours, neurological disorders, developmental abnormalities and foetal
CC   deficiencies, blood disorders, leukaemias, diseases of the immune system
CC   (including allergies or asthma), hepatic disease, Alzheimer's and
CC   cognitive disorders, schizophrenia, prostate diseases, autoimmune
CC   disorders and AIDS. The polypeptides are also useful for identifying
CC   their binding partners.
XX
SQ   Sequence 2017 BP; 572 A; 406 C; 470 G; 569 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 20;  Length 2017;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4861 GCAAACATCAAATAATATAA 4880
             ||||||||||||||||||||
Db       1612 GCAAACATCAAATAATATAA 1593


RESULT 53
AAZ56764/c
ID   AAZ56764 standard; cDNA; 2031 BP.
XX
AC   AAZ56764;
XX
DT   23-MAR-2000  (first entry)
XX
DE   Human transmembrane protein HTMPN-67 encoding cDNA.
XX
KW   Human; transmembrane protein; HTMPN; diagnosis; immunospecific;
KW   antiproliferative; neuroprotective; immune disorder;
KW   reproductive disorder; smooth muscle disorder; neurological disorder;
KW   gastrointestinal disorder; developmental disorder;
KW   cell proliferative disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO9961471-A2.
XX
PD   02-DEC-1999.
XX
PF   28-MAY-1999;   99WO-US11904.
XX
PR   29-MAY-1998;   98US-0087260.
PR   02-JUL-1998;   98US-0091674.
PR   02-OCT-1998;   98US-0102954.
PR   24-NOV-1998;   98US-0109869.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Tang YT,  Lal P,  Hillman JL,  Yue H,  Guegler KJ,  Corley NC;
PI   Bandman O,  Patterson C,  Gorgone GA,  Kaser MR,  Baughn MR;
PI   Au-Young J;
XX
DR   WPI; 2000-072605/06.
DR   P-PSDB; AAY57943.
XX
PT   Proteins, polynucleotides, vectors, host cells and antibodies used to
PT   diagnose, treat or prevent immune, reproductive, smooth muscle,
PT   neurological, gastrointestinal, developmental and cell proliferative
PT   disorders -
XX
PS   Claim 9; Page 220; 229pp; English.
XX
CC   AAZ56698 to AAZ56776 encode AAY57877 to AAY57955 which represent human
CC   transmembrane proteins designated HTMPN-1 to HTMPN-79, respectively.
CC   The transmembrane protein have immunospecific, antiproliferative and
CC   neuroprotective activities. The human transmembrane proteins,
CC   polynucleotides encoding them and other compositions and methods from
CC   the present invention, can be used for the diagnosis, treatment or
CC   prevention of immune, reproductive, smooth muscle, neurological,
CC   gastrointestinal, developmental and cell proliferative disorders. The
CC   HTMPN's can be used to treat or prevent disorders associated with a
CC   decreased expression or activity of HTMPN.
XX
SQ   Sequence 2031 BP; 534 A; 428 C; 495 G; 574 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 21;  Length 2031;
```

Best Local Similarity 100.0%; Pred. No. 89;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       4861 GCAAACATCAAATAATATAA 4880
              ||||||||||||||||||||
Db       1675 GCAAACATCAAATAATATAA 1656
```

RESULT 54
AAK51631/c
ID AAK51631 standard; cDNA; 2337 BP.
XX
AC AAK51631;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 176.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR P-PSDB; AAM78498.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 1; Page 926-928; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 2337 BP; 587 A; 558 C; 564 G; 628 T; 0 other;

Query Match 0.2%; Score 20; DB 22; Length 2337;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       4861 GCAAACATCAAATAATATAA 4880
              ||||||||||||||||||||
Db       1968 GCAAACATCAAATAATATAA 1949
```

RESULT 55
ABK28365/c
ID ABK28365 standard; DNA; 9238 BP.
XX
AC ABK28365;
XX
DT 23-APR-2002 (first entry)
XX
DE DNA transcription associated genomic DNA #120.
XX
KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW immunological disorder; Werner syndrome; developmental disorder;
KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW polyglutamine disorder; solid tumour.
XX
OS Unidentified.
XX
PN WO200192565-A2.
XX
PD 06-DEC-2001.
XX
PF 06-APR-2001; 2001WO-EP03973.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-090046/12.
XX
PT New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological
PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid
PT tumours or cancer -
XX
PS Claim 1; SEQ ID No 239; 32pp; English.
XX
CC The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention.

```
CC   Note; The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from the
CC   European Patent Office.
XX
SQ   Sequence 9238 BP; 2692 A; 159 C; 1985 G; 4402 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 24;  Length 9238;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4492 TACAAAATCCTTATAAAAAA 4511
             ||||||||||||||||||||
Db       6908 TACAAAATCCTTATAAAAAA 6889


RESULT 56
ABL92227/c
ID   ABL92227 standard; DNA; 11131 BP.
XX
AC   ABL92227;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Chemically treated DNA repair gene fragment complementary to#18.
XX
KW   DNA repair; cytosine methylation; PMS2L1; PMS2L12; PMS2L2; PMS2L3;
KW   PMS2; L4; PMS2L5; PMS2L6; MGMT; MSH2; NUDT1; TDG; INPPL1; RFC4;
KW   DDIT1L; FANCB; XRCC8; ataxia telangiectasia; aging; Bloom's syndrome;
KW   Cockayne syndrome; Nijmegen breakage syndrome; Werner syndrome;
KW   immunodeficiency; trichthiodystrophy; Fanconi's anaemia; solid tumour;
KW   cancer; ds.
XX
OS   Unidentified.
XX
PN   WO200181622-A2.
XX
PD   01-NOV-2001.
XX
PF   06-APR-2001; 2001WO-EP03972.
XX
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-034446/04.
XX
PT   New nucleic acid derived from genes associated with DNA repair, useful
PT   for diagnosis, e.g. of ataxia telangiectasia, by determination of
PT   cytosine methylation -
XX
PS   Claim 1; SEQ ID NO 36; 25pp + sequence listing; English.
XX
CC   The invention relates to nucleic acids containing a sequence of at least
CC   18 nucleotides of chemically treated DNA of genes associated with DNA
CC   repair, and their complements. The invention also relates to nucleic
CC   acids comprising at least 18 base pairs of the chemically pretreated DNA
CC   of genes associated with DNA repair selected from PMS2L1, PMS2L12,
CC   PMS2L2, PMS2L3, PMS2, L4, PMS2L5, PMS2L6, MGMT, MSH2, NUDT1, TDG, INPPL1,
CC   RFC4, DDIT1L, FANCB, or XRCC8. Nucleic acids of the invention and related
CC   oligomers, are useful for diagnosis of diseases associated with gene
CC   repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,
CC   Cockayne syndrome, Nijmegen breakage syndrome or Werner syndrome,
CC   immunodeficiency, trichthiodystrophy, Fanconi's anaemia, solid tumours
CC   and cancer, particularly by determining status of cytosine methylation
CC   and/or by detecting single-nucleotide polymorphisms. Determination of
CC   individual methylation patterns may allow development of individualised
CC   therapies. The sequences given in records ABL92192-ABL92335 represent
CC   chemically pre-treated DNA fragments from genes associated with DNA
CC   repair, and their complements.
CC   Note: The sequence data for this patent is not represented in the
CC   specification, but is based on sequence information supplied by the
CC   European Patent Office.
XX
SQ   Sequence 11131 BP; 2874 A; 328 C; 2644 G; 5285 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 24;  Length 11131;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2364 CATCTACTAAATAAATAAAA 2383
             ||||||||||||||||||||
Db       9539 CATCTACTAAATAAATAAAA 9520


RESULT 57
ABL49328/c
ID   ABL49328 standard; DNA; 11131 BP.
XX
AC   ABL49328;
XX
DT   01-MAY-2002  (first entry)
XX
DE   Human polynucleotide associated with DNA replication SEQ ID NO 28.
XX
KW   Human; cytostatic; neuroprotective; nootropic; immunostimulant;
KW   gene therapy; gene regulation; DNA replication; CENPB; DNA2L; ATR; CHD1L;
KW   ERCC3; SNRPA1; RAD50; LIG2; cytosine methylation; Ataxia telangiectasia;
KW   ATR-X; Bloom's syndrome; tumour; cancer; methylation; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200177377-A2.
XX
PD   18-OCT-2001.
XX
PF   06-APR-2001; 2001WO-EP03971.
XX
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-017471/02.
XX
PT   New nucleic acid sequences from chemically modified genes associated
PT   with DNA replication, useful for analysing cytosine methylations for
PT   diagnosis and therapy of diseases e.g. Ataxia telangiectasia  -
XX
PS   Claim 1; SEQ ID NO 28; 23pp + Sequence Listing; English.
XX
CC   The invention relates to nucleic acid sequences comprising at least 18
CC   bases of a chemically pretreated gene associated with gene regulation,
CC   selected from 94 genes (ABL49301-ABL49394) and/or complementary sequences
CC   associated with DNA replication, CENPB, DNA2L, ATR, CHD1L, ERCC3, SNRPA1,
CC   RAD50 and LIG2. The chemical pretreatment converts cytosine bases
CC   unmethylated at the 5-position to uracil or another base with
CC   hybridisation behaviour dissimilar to cytosine, to enable analysis of
CC   cytosine methylations. The DNA sequences and method are useful in the
CC   diagnosis of diseases (or predisposition to diseases) associated with DNA
CC   replication and in therapy of such diseases, by enabling analysis of the
CC   cytosine methylation patterns of such genes. They are especially useful
CC   in diagnosis and therapy of e.g. Ataxia telangiectasia, ATR-X, Bloom's
CC   syndrome, solid tumours and cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification and was supposed to be available directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences. However, the sequence data did
CC   not correspond to that referred to in the specification. The present data
```

```
CC   is taken from EPO data for the patent.
XX
SQ.  Sequence 11131 BP; 2874 A; 328 C; 2644 G; 5285 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 24;  Length 11131;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2364 CATCTACTAAATAAATAAAA 2383
             ||||||||||||||||||||
Db       9539 CATCTACTAAATAAATAAAA 9520


RESULT 58
ABL32753/c
ID   ABL32753 standard; DNA; 11131 BP.
XX
AC   ABL32753;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 726.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 726; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 11131 BP; 2874 A; 328 C; 2644 G; 5285 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 24;  Length 11131;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2364 CATCTACTAAATAAATAAAA 2383
             ||||||||||||||||||||
Db       9539 CATCTACTAAATAAATAAAA 9520


RESULT 59
AAS46431/c
ID   AAS46431 standard; DNA; 11327 BP.
XX
AC   AAS46431;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #153.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW   cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP02955.
XX
PR   15-MAR-2000; 2000DE-1013847.
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumour suppressor
PT   genes and oncogenes, useful in designing primers and probes for
PT   analysing diseases associated with cytosine methylation state e.g.
PT   cancer  -
XX
PS   Claim 1; SEQ ID No 153; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and
CC   oncogenes having a sequence taken from 536 (actually 533 since
CC   numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC   (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC   peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC   form part of a set of probes for detecting the cytosine methylation state
CC   and/or single nucleotide polymorphisms and also to be used in an
CC   array for analysing diseases associated with CpG dinucleotides e.g.
CC   cancers and tumours. The probes can also be used in a method for
CC   ascertaining genetic and/or epigenetic parameters for the diagnosis
CC   and/or therapy of existing diseases or the predisposition to specific
CC   diseases, by analysing cytosine methylations. The parameters may be
CC   compared to another set of genetic and/or epigenetic parameters, the
CC   differences serving as basis for diagnosis and/or prognosis events which
CC   are disadvantageous to patients. The present sequence is one of the
CC   533 genomic sequences derived from tumour suppressor genes and
CC   oncogenes.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 11327 BP; 2724 A; 316 C; 3043 G; 5244 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 22;  Length 11327;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       6775 AAAAAATTTTTTAACCAAAC 6794
              ||||||||||||||||||||
Db       3047 AAAAAATTTTTTAACCAAAC 3028


RESULT 60
ABL56203
ID   ABL56203 standard; DNA; 32392 BP.
XX
AC   ABL56203;
XX
DT   01-JUL-2002  (first entry)
XX
DE   AmEPV genome fragment#5.
XX
KW   AmEPV; gene therapy; viral vector; chromosome mapping; gene mapping;
KW   genetic deficiency disorder; ds.
XX
OS   Amsacta moorei entomopoxvirus.
XX
PN   WO200212526-A2.
XX
PD   14-FEB-2002.
XX
PF   10-AUG-2001; 2001WO-US25287.
XX
PR   10-AUG-2000; 2000US-224479P.
PR   14-SEP-2000; 2000US-0662254.
XX
PA   (UYFL ) UNIV FLORIDA.
XX
PI   Moyer RW,  Li Y,  Bawden AL;
XX
DR   WPI; 2002-227161/28.
XX
PT   Novel recombinant entomopox virus vector useful for delivering
PT   polynucleotide encoding protein to vertebrate cell, comprises
PT   polynucleotide encoding protein operably linked with heterologous
PT   promoter sequence  -
XX
PS   Disclosure; Page 226-242; 326pp; English.
XX
CC   The invention relates to a recombinant entomopox virus (EPV) vector,
CC   comprising a polynucleotide encoding a protein operably linked with a
CC   heterologous promoter sequence. The invention also concerns methods for
CC   providing gene therapy for genetic deficiency disorders. Vectors of the
CC   invention are useful for delivering a polynucleotide encoding a protein
CC   to a vertebrate cell preferably a mammalian cell, such as a human cell.
CC   The vector is introduced into the vertebrate cell by infection in a viral
CC   particle, or by transfection, transduction, or injection either in vitro
CC   or in vivo. The vector is useful for the delivery and expression of
CC   biologically useful proteins in gene therapy protocols, and for
CC   delivering large DNA segments for engineering of vertebrate cells.
CC   Polynucleotides of the invention have applications in techniques such as
CC   their use as insertion sites for foreign genes of interest, hybridisation
CC   probes, for chromosome and gene mapping, in PCR technologies, and in the
CC   production of sense or antisense nucleic acids. Vectors of the invention
CC   provide for stable integration and expression of heterologous DNA in host
CC   cells, and are adapted for accepting large heterologous polynucleotide
CC   inserts which can be delivered in an infected or transformed cell and
CC   expressed in a stable fraction. The current sequence represents a
CC   fragment of the genome of the genus B entomopoxvirus from amsacta moorei
CC   (AmEPV).
XX
SQ   Sequence 32392 BP; 13748 A; 2577 C; 2550 G; 13517 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 24;  Length 32392;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4801 TATATAATATGTATAGAATT 4820
              ||||||||||||||||||||
Db       4744 TATATAATATGTATAGAATT 4763


RESULT 61
AAV73803
ID   AAV73803 standard; DNA; 35100 BP.
XX
AC   AAV73803;
XX
DT   25-FEB-1999  (first entry)
XX
DE   KSHV LUR DNA (nucleotides 35,101-70,200).
XX
KW   Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHFR; Bcl-2;
KW   dihydrofolate reductase; LUR; long unique region; vaccine; prophylaxis;
KW   diagnosis; treatment; HHV8; thymidine kinase; glycoprotein H;
KW   capsid protein; packaging protein; helicase primase; glycoprotein M;
KW   viral protein kinase; alkaline exonuclease; virion assembly protein;
KW   uracil DNA glycosylase; UDG; glycoprotein L; ds.
XX
OS   Kaposi's sarcoma-associated herpesvirus.
XX
PN   US5849564-A.
XX
PD   15-DEC-1998.
XX
PF   29-NOV-1996;   96US-0770379.
XX
PR   29-NOV-1996;   96US-0770379.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Bohenzky RA,  Chang Y,  Edelman IS,  Moore PS,  Russo JJ;
XX
DR   WPI; 1999-069741/06.
XX
PT   Kaposi's sarcoma-associated herpes virus nucleic acid - encodes
PT   di:hydro:folate reductase and is useful for treatment, prophylaxis
PT   or diagnosis of Kaposi's sarcoma
XX
PS   Disclosure; Column 97-126; 109pp; English.
XX
CC   This sequence is a fragment of the Kaposi's sarcoma-associated
CC   herpesvirus (KSHV) LUR (long unique region). This fragment contains
CC   coding regions for ORF20, OFR21 which encodes thymidine kinase, ORF22
CC   which encodes glycoprotein H, ORF23, ORF24, ORF25 which encodes the
CC   major capsid protein, ORF26 which encodes capsid protein II, ORF27,
CC   ORF28, ORF29b which encodes packaging protein II, ORF30, ORF31, ORF32,
CC   ORF33, ORF29a which encodes packaging protein I, ORF34, ORF35, ORF36
CC   which encodes viral protein kinase, ORF37 which encodes alkaline
CC   exonuclease, ORF38, ORF39 which encodes glycoprotein M, ORF40 which
CC   encodes helicase primase subunit I, ORF41 which encodes helicase primase
CC   subunit II, ORF42, ORF43 which encodes capsid protein III, ORF44 which
CC   encodes helicase primase subunit III, ORF45 which encodes the virion
CC   assembly protein, ORF46 which encodes uracil DNA glycosylase (UDG),
CC   ORF47 which encodes glycoprotein L. KSHV is a new human Herpesvirus
CC   (HHV8) believed to cause Kaposi's sarcoma (KS) which is the most common
CC   form of neoplasm occurring in persons with acquired immune deficiency
CC   syndrome (AIDS). The DHFR protein is useful for vaccination, prophylaxis,
CC   diagnosis and treatment of a subject with Kaposi's sarcoma and for
CC   detecting expression of a DNA virus associated with Kaposi's sarcoma in a
CC   cell.
XX
SQ   Sequence 35100 BP; 8016 A; 9231 C; 9477 G; 8376 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 20;  Length 35100;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3776 GTGAACCAATCCCAGCCAAG 3795
              ||||||||||||||||||||
Db      33802 GTGAACCAATCCCAGCCAAG 33821
```

```
RESULT 62
AAV19941
ID   AAV19941 standard; DNA; 137507 BP.
XX
AC   AAV19941;
XX
DT   03-AUG-1998  (first entry)
XX
DE   KSHV long unique coding region and terminal repeat.
XX
KW   KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II;
KW   interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;
KW   complement-binding protein; glycoprotein; capsid protein IV; infection;
KW   immediate early protein; Kaposi's sarcoma; protective vaccine; lymphoma;
KW   lymphoproliferative disease; leukaemia; splenomegaly; mycosis fungoides;
KW   HIV immune status; anti-inflammatory agent; therapy; ds.
XX
OS   Kaposi's sarcoma-associated herpes virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1142..2794
FT                   /*tag= a
FT                   /product= complement-binding protein
FT   CDS             8699..11236
FT                   /*tag= b
FT                   /product= glycoprotein B
FT   CDS             complement (17261..17875)
FT                   /*tag= c
FT                   /product= interleukin 6
FT   CDS             complement (21548..21832)
FT                   /*tag= d
FT                   /product= macrophage inflammatory protein II
FT   CDS             complement (27137..27424)
FT                   /*tag= e
FT                   /product= interferon regulatory factor 1
FT   CDS             28661..29741
FT                   /*tag= f
FT                   /product= protein T1.1
FT   CDS             complement (58976..60175)
FT                   /*tag= g
FT                   /product= glycoprotein M
FT   CDS             complement (69412..69915)
FT                   /*tag= h
FT                   /product= glycoprotein L
FT   CDS             complement (88410..88910)
FT                   /*tag= i
FT                   /product= interferon regulatory factor 2
FT   CDS             89600..90541
FT                   /*tag= j
FT                   /product= interferon regulatory factor 3
FT   CDS             90173..90643
FT                   /*tag= k
FT                   /product= glycoprotein X
FT   CDS             complement (93636..94127)
FT                   /*tag= l
FT                   /product= interferon regulatory factor 4
FT   CDS             complement (111931..112443)
FT                   /*tag= m
FT                   /product= capsid protein IV
FT   CDS             complement (123808..127296)
FT                   /*tag= n
FT                   /product= immediate early protein
XX
PN   WO9804576-A1.
XX
PD   05-FEB-1998.
XX
PF   22-JUL-1997;    97WO-US13346.
XX
PR   29-NOV-1996;    96US-0757669.
PR   25-JUL-1996;    96US-0686243.
PR   25-JUL-1996;    96US-0686349.
PR   25-JUL-1996;    96US-0686350.
PR   25-JUL-1996;    96US-0687253.
PR   25-JUL-1996;    96US-0688814.
PR   05-SEP-1996;    96US-0708678.
PR   10-OCT-1996;    96US-0728323.
PR   13-NOV-1996;    96US-0747887.
PR   13-NOV-1996;    96US-0748640.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Bohenzky RA,  Chang Y,  Edelman IS,  Moore PS,  Russo JJ;
XX
DR   WPI; 1998-130615/12.
XX
PT   New nucleic acid encoding Kaposi's sarcoma associated herpes virus
PT   proteins - useful for, e.g. detecting levels of HHV8 in, and
PT   preparation of vaccines for treatment of, HIV patients
XX
PS   Example 2; Page 135-203; 230pp; English.
XX
CC   This sequence represents the long unique region and terminal repeat of
CC   the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known
CC   as human herpes virus 8 (HHV8). This sequence contains the DNAs of the
CC   invention which encode KSHV polypeptides selected from: (a) viral
CC   macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);
CC   (c) viral IRF 1; (d) complement-binding protein; glycoproteins B, M or L;
CC   (d) capsid protein IV encoded by ORF65; and (e) immediate early protein
CC   encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded
CC   by it, and antibodies (Ab) specific for the proteins are useful for
CC   detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body
CC   fluids or tissue samples. HHV8 infections can be treated with antisense
CC   or triplex forming molecules or agents that bind specifically to the
CC   protein. Ab may be used for prophylaxis or treatment of HHV8 infection,
CC   while the protein can be used in protective vaccines. Ab may also be used
CC   to differentiate between lymphomas, and HHV8 may be implicated in many
CC   other lymphoproliferative diseases such as lymphomas, leukaemia,
CC   splenomegaly and mycosis fungoides. Cells and animals containing the
CC   nucleic acid are useful for drug screening. HHV8-derived peptides can be
CC   used as targets for antiviral drugs, e.g. dihydrofolate reductase gene
CC   can be inhibited with methotrexate. These can also be used to determine
CC   the immune status of a patient infected with HIV. HHV8 derived protein
CC   viral MIP III may be used as an anti-inflammatory agent for,
CC   e.g. treating rheumatoid arthritis. This sequence is stated as containing
CC   81 open reading frames.
XX
SQ   Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T; 0 other;

  Query Match             0.2%;  Score 20;  DB 19;  Length 137507;
  Best Local Similarity   100.0%;  Pred. No. 79;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3776 GTGAACCAATCCCAGCCAAG 3795
              ||||||||||||||||||||
Db      68902 GTGAACCAATCCCAGCCAAG 68921


RESULT 63
AAX91990
ID   AAX91990 standard; DNA; 1230025 BP.
XX
AC   AAX91990;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Nucleotide sequence of the complete genome of Chlamydia pneumoniae.
XX
KW   Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW   sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW   vaccine; neutralising epitope; ss.
XX
OS   Chlamydia pneumoniae.
XX
PN   WO9927105-A2.
XX
```

```
PD   03-JUN-1999.
XX
PF   20-NOV-1998;   98WO-IB01890.
XX
PR   04-NOV-1998;   98US-0107078.
PR   21-NOV-1997;   97FR-0014673.
XX
PA   (GEST ) GENSET.
XX
PI   Griffais R;
XX
DR   WPI; 1999-357842/30.
XX
PT   Genome sequence of Chlamydia pneumoniae
XX
PS   Claim 1; Page 291-611; 1912pp; English.
XX
CC   The present sequence represents the complete genome of Chlamydia
CC   pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes
CC   respiratory disease such as pneumonia and bronchitis and is thought
CC   to be a contributing factor in heart disease, sarcoidosis, sinusitis,
CC   purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
CC   encoded by the open reading frames of the C. pneumoniae genome (see
CC   AAY34584-Y35879) can be used in immunogenic compositions as vaccines.
CC   Vectors containing C. pneumoniae nucleotides sequences can also be
CC   used as immunogenic compositions, especially where the vector directs
CC   the expression of a neutralising epitope of C. pneumoniae.
XX
SQ   Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

  Query Match             0.2%;  Score 20;  DB 20;  Length 1230025;
  Best Local Similarity   100.0%;  Pred. No. 73;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6172 TTAACACCCTTAACTGGAAT 6191
             ||||||||||||||||||||
Db     907689 TTAACACCCTTAACTGGAAT 907708


RESULT 64
ABX40498/c
ID   ABX40498 standard; cDNA; 221 BP.
XX
AC   ABX40498;
XX
DT   20-FEB-2003  (first entry)
XX
DE   Bovine EST associated with lactation/muscle/fat deposition #5663.
XX
KW   Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW   muscle deposition; fat deposition; genome mapping; gene identification;
KW   gene analysis; cattle breeding.
XX
OS   Bos Taurus.
XX
PN   US2002137139-A1.
XX
PD   26-SEP-2002.
XX
PF   24-SEP-2001; 2001US-0960352.
XX
PR   12-JAN-1999;   99US-115707P.
PR   11-JAN-2000; 2000US-0480902.
XX
PA   (BYAT/) BYATT J C.
PA   (MATH/) MATHIALAGAN N.
PA   (TAON/) TAO N.
PA   (WARR/) WARREN W C.
XX
PI   Byatt JC,  Mathialagan N,  Tao N,  Warren WC;
XX
DR   WPI; 2003-110599/10.
XX
PT   New nucleic acid associated with lactation, and muscle and fat
PT   deposition, useful for genome mapping, gene identification and
PT   analysis, cattle breeding, or for genetically improving cattle
XX
PS   Claim 2; SEQ ID No 5663; 245pp; English.
XX
CC   The invention relates to a purified nucleic acid molecule associated with
CC   lactation or muscle and fat deposition (designated LMFD), derived
CC   from cattle, and the LMFD nucleic acid can specifically hybridise to a
CC   second nucleic acid molecule comprising any of 15112 nucleotide
CC   sequences, appearing as ABX34836-ABX49947, or complements of them.
CC   Also included are; (1) a transformed cell having a nucleic acid
CC   comprising an LMFD nucleic acid linked to a promoter and a 3' non-
CC   translated sequence that functions in the cell to cause termination of
CC   transcription and addition of polyadenylated ribonucleotides to a 3' end
CC   of the mRNA molecule; and (2) determining a level or pattern of a
CC   molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC   nucleic acid (comprising any of the 15112 nucleic acid sequences or its
CC   complement or fragment) with a complementary nucleic acid molecule
CC   obtained from the bovine cell or tissue, where hybridisation between the
CC   marker nucleic acid and the complementary nucleic acid permits the
CC   detection of the molecule; and (b) detecting the level or pattern of the
CC   complementary nucleic acid, where the detection of the complementary
CC   nucleic acid is predictive of the level or pattern of the molecule.
CC   The LMFD nucleic acid is used for determining a level or pattern
CC   of a molecule in a bovine cell or tissue. It is useful for genome
CC   mapping, gene identification and analysis, cattle breeding, preparation
CC   of constructs for use in cattle gene expression, or for genetically
CC   improving cattle. The present sequence is one of the 15112 bovine
CC   LMFD EST (expressed sequence tag) nucleic acids.
CC   Note: The present sequence was not shown in the specification but
CC   was obtained in electronic format from the USPTO web site:
CC   seqdata.uspto.gov/sequence.html?DocID=20020137139.
XX
SQ   Sequence 221 BP; 58 A; 49 C; 41 G; 73 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 25;  Length 221;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6207 ACCGTGAACTTCCTGATGT 6225
             |||||||||||||||||||
Db        216 ACCGTGAACTTCCTGATGT 198


RESULT 65
ABL84861
ID   ABL84861 standard; cDNA; 268 BP.
XX
AC   ABL84861;
XX
DT   17-MAY-2002  (first entry)
XX
DE   Human ovarian cancer related cDNA clone SEQ ID NO:7839.
XX
KW   Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200192581-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US17756.
XX
PR   26-MAY-2000; 2000US-207484P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Algate PA,  Harlocker SL,  Jones R;
XX
DR   WPI; 2002-122075/16.
XX
```

PT Composition for therapy and diagnosis of ovarian cancer comprising
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT polypeptide, antibody specific to polypeptide or T cell expressing
PT polypeptide -
XX
PS Claim 1; SEQ ID 7839; 489pp; English.
XX
CC The present invention describes a composition (I) comprising: carriers
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence
CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to
CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell
CC population of (II), or antigen presenting cells that express (II).
CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to
CC (S1) can be used for detecting ovarian cancer in a patient's biological
CC sample preferably serum or ovarian tissue. The method comprises
CC contacting a biological sample from a patient with (IV), detecting the
CC amount of polynucleotide hybridising to (IV) and comparing the amount to
CC a predetermined cutoff value and thereby detecting ovarian cancer in the
CC patient, where the amount of polynucleotide hybridising to (IV) is
CC detected preferably by polymerase chain reaction (PCR). (I) comprising
CC (III) and/or (II) is useful for stimulating and/or expanding T cells
CC specific for an ovarian tumour protein comprising contacting T cells
CC with (III) or (II). (III) is useful in design and preparation of
CC ribozyme molecules for inhibiting expression of the tumour polypeptides
CC and proteins in tumour cells; and to isolate a full length gene from a
CC suitable library e.g., a tumour cDNA library using well known
CC techniques.
XX
SQ Sequence 268 BP; 97 A; 46 C; 41 G; 84 T; 0 other;

Query Match 0.2%; Score 19; DB 24; Length 268;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4715 TTAAAAAAAATCCACATGT 4733
||||||||||||||||||
Db 111 TTAAAAAAAATCCACATGT 129

RESULT 66
ABX86121
ID ABX86121 standard; cDNA; 284 BP.
XX
AC ABX86121;
XX
DT 24-APR-2003 (first entry)
XX
DE Corn ear-derived polynucleotide (cpd) #4581.
XX
KW Corn ear-derived polynucleotide; cdp; cDNA library; SATMON022;
KW SATMON023; structural gene; functional gene; regulatory gene;
KW corn ear-specific profile; gene transcription; gene expression;
KW hybrid plant; desirable trait expression; plant breeding program;
KW inheritance; desired characteristic; growth; development;
KW disease resistance; environmental adaptability; quality; yield;
KW multigene trait; plant; gene; ss.
XX
OS Zea mays.
XX
PN US6476212-B1.
XX
PD 05-NOV-2002.
XX
PF 14-MAY-1999; 99US-0313294.
XX
PR 26-MAY-1998; 98US-086722P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Lalgudi RV, Ito LY, Sherman BK;
XX
DR WPI; 2003-208840/20.
XX
PT Novel purified corn-ear derived polynucleotide useful as hybridization
PT probe for detecting polynucleotide in sample, and for identifying,
PT evaluating, and altering desired characteristics associated with
PT growth, development -
XX
PS Examples; SEQ ID No 4581; 390pp; English.
XX
CC The present invention relates to the isolation of corn ear-derived
CC polynucleotides (cdps) from the corn (Zea mays) cDNA libraries
CC SATMON022 and SATMON023. Some of the cpds uniquely identify structural,
CC functional, and regulatory genes of corn ear. The polynucleotides
CC sequences are useful for detecting cpds in a sample, for producing
CC a corn ear-specific profile of gene transcription, for detecting
CC altered gene expression in inbred or hybrid plants, and for screening
CC several molecules for specific binding to the polynucleotide. The cdps
CC are useful to identify, isolate, or extend identical or related
CC corn-ear nucleic acid sequences from DNA libraries, and in nucleic
CC acid amplification or hybridisation techniques to follow the
CC expression of desirable traits through plant breeding programs.
CC Preferably, the cdps are used to identify, evaluate, alter, or
CC follow the inheritance of desired characteristics associated with
CC growth and development, disease resistance, environmental adaptability,
CC quality, and yield of corn. The cdps are also useful as molecular
CC markers for studying inheritance and multigene traits in a plant
CC breeding program. The cpds are useful for producing purified corn-ear
CC polypeptides by recombinant techniques. They are also useful in
CC diagnostic assays to detect or confirm conditions or diseases
CC associated with abnormal levels of cdp expression. ABX81541-ABX89140
CC represent corn ear-derived polynucleotides (cpds) of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site at seqdata.uspto.gov/psipsDIDEntry.html.
XX
SQ Sequence 284 BP; 67 A; 64 C; 79 G; 73 T; 1 other;

Query Match 0.2%; Score 19; DB 25; Length 284;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8700 TGGCGGCCGGCTCTCCGCC 8718
||||||||||||||||||
Db 22 TGGCGGCCGGCTCTCCGCC 40

RESULT 67
ABX81628
ID ABX81628 standard; cDNA; 292 BP.
XX
AC ABX81628;
XX
DT 24-APR-2003 (first entry)
XX
DE Corn ear-derived polynucleotide (cpd) #88.
XX
KW Corn ear-derived polynucleotide; cdp; cDNA library; SATMON022;
KW SATMON023; structural gene; functional gene; regulatory gene;
KW corn ear-specific profile; gene transcription; gene expression;
KW hybrid plant; desirable trait expression; plant breeding program;
KW inheritance; desired characteristic; growth; development;
KW disease resistance; environmental adaptability; quality; yield;
KW multigene trait; plant; gene; ss.
XX
OS Zea mays.
XX
PN US6476212-B1.
XX
PD 05-NOV-2002.
XX
PF 14-MAY-1999; 99US-0313294.
XX
PR 26-MAY-1998; 98US-086722P.
XX

PA (INCY-) INCYTE GENOMICS INC.
XX
PI Lalgudi RV, Ito LY, Sherman BK;
XX
DR WPI; 2003-208840/20.
XX
PT Novel purified corn-ear derived polynucleotide useful as hybridization
PT probe for detecting polynucleotide in sample, and for identifying,
PT evaluating, and altering desired characteristics associated with
PT growth, development
XX
PS Examples; SEQ ID No 88; 390pp; English.
XX
CC The present invention relates to the isolation of corn ear-derived
CC polynucleotides (cdps) from the corn (Zea mays) cDNA libraries
CC SATMON022 and SATMON023. Some of the cpds uniquely identify structural,
CC functional, and regulatory genes of corn ear. The polynucleotides
CC sequences are useful for detecting cpds in a sample, for producing
CC a corn ear-specific profile of gene transcription, for detecting
CC altered gene expression in inbred or hybrid plants, and for screening
CC several molecules for specific binding to the polynucleotide. The cdps
CC are useful to identify, isolate, or extend identical or related
CC corn-ear nucleic acid sequences from DNA libraries, and in nucleic
CC acid amplification or hybridisation techniques to follow the
CC expression of desirable traits through plant breeding programs.
CC Preferably, the cdps are used to identify, evaluate, alter, or
CC follow the inheritance of desired characteristics associated with
CC growth and development, disease resistance, environmental adaptability,
CC quality, and yield of corn. The cdps are also useful as molecular
CC markers for studying inheritance and multigene traits in a plant
CC breeding program. The cpds are useful for producing purified corn-ear
CC polypeptides by recombinant techniques. They are also useful in
CC diagnostic assays to detect or confirm conditions or diseases
CC associated with abnormal levels of cdp expression. ABX81541-ABX89140
CC represent corn ear-derived polynucleotides (cpds) of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site at seqdata.uspto.gov/psipsDIDEntry.html.
XX
SQ Sequence 292 BP; 71 A; 63 C; 80 G; 74 T; 4 other;

Query Match 0.2%; Score 19; DB 25; Length 292;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8700 TGGCGGCCGGCTCTCCGCC 8718
||||||||||||||||||||
Db 24 TGGCGGCCGGCTCTCCGCC 42

RESULT 68
AAK89762
ID AAK89762 standard; DNA; 305 BP.
XX
AC AAK89762;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen genomic sequence SEQ ID NO: 3338.
XX
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX
OS Homo sapiens.
XX
PN WO200155314-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01324.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases -
XX
PS Disclosure; SEQ ID NO 3338; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention.
XX
SQ Sequence 305 BP; 97 A; 52 C; 85 G; 71 T; 0 other;

Query Match 0.2%; Score 19; DB 22; Length 305;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7578 ATCTCTTTAACCTGTTCTG 7596
||||||||||||||||||||
Db 287 ATCTCTTTAACCTGTTCTG 305

RESULT 69
AAK89766
ID AAK89766 standard; DNA; 305 BP.
XX
AC AAK89766;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen genomic sequence SEQ ID NO: 3342.
XX
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX
OS Homo sapiens.
XX
PN WO200155314-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01324.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.

```
PR  18-AUG-2000; 2000US-0226279.
PR  22-AUG-2000; 2000US-0226681.
PR  22-AUG-2000; 2000US-0226868.
PR  22-AUG-2000; 2000US-0227182.
PR  23-AUG-2000; 2000US-0227009.
PR  30-AUG-2000; 2000US-0228924.
PR  01-SEP-2000; 2000US-0229287.
PR  01-SEP-2000; 2000US-0229343.
PR  01-SEP-2000; 2000US-0229344.
PR  01-SEP-2000; 2000US-0229345.
PR  05-SEP-2000; 2000US-0229509.
PR  05-SEP-2000; 2000US-0229513.
PR  06-SEP-2000; 2000US-0230437.
PR  06-SEP-2000; 2000US-0230438.
PR  08-SEP-2000; 2000US-0231242.
PR  08-SEP-2000; 2000US-0231243.
PR  08-SEP-2000; 2000US-0231244.
PR  08-SEP-2000; 2000US-0231413.
PR  08-SEP-2000; 2000US-0231414.
PR  08-SEP-2000; 2000US-0232080.
PR  08-SEP-2000; 2000US-0232081.
PR  12-SEP-2000; 2000US-0231968.
PR  14-SEP-2000; 2000US-0232397.
PR  14-SEP-2000; 2000US-0232398.
PR  14-SEP-2000; 2000US-0232399.
PR  14-SEP-2000; 2000US-0232400.
PR  14-SEP-2000; 2000US-0232401.
PR  14-SEP-2000; 2000US-0233063.
PR  14-SEP-2000; 2000US-0233064.
PR  14-SEP-2000; 2000US-0233065.
PR  21-SEP-2000; 2000US-0234223.
PR  21-SEP-2000; 2000US-0234274.
PR  25-SEP-2000; 2000US-0234997.
PR  25-SEP-2000; 2000US-0234998.
PR  26-SEP-2000; 2000US-0235484.
PR  27-SEP-2000; 2000US-0235834.
PR  27-SEP-2000; 2000US-0235836.
PR  29-SEP-2000; 2000US-0236327.
PR  29-SEP-2000; 2000US-0236367.
PR  29-SEP-2000; 2000US-0236368.
PR  29-SEP-2000; 2000US-0236369.
PR  29-SEP-2000; 2000US-0236370.
PR  02-OCT-2000; 2000US-0236802.
PR  02-OCT-2000; 2000US-0237037.
PR  02-OCT-2000; 2000US-0237038.
PR  02-OCT-2000; 2000US-0237039.
PR  02-OCT-2000; 2000US-0237040.
PR  13-OCT-2000; 2000US-0239935.
PR  13-OCT-2000; 2000US-0239937.
PR  20-OCT-2000; 2000US-0240960.
PR  20-OCT-2000; 2000US-0241221.
PR  20-OCT-2000; 2000US-0241785.
PR  20-OCT-2000; 2000US-0241786.
PR  20-OCT-2000; 2000US-0241787.
PR  20-OCT-2000; 2000US-0241808.
PR  20-OCT-2000; 2000US-0241809.
PR  20-OCT-2000; 2000US-0241826.
PR  01-NOV-2000; 2000US-0244617.
PR  08-NOV-2000; 2000US-0246474.
PR  08-NOV-2000; 2000US-0246475.
PR  08-NOV-2000; 2000US-0246476.
PR  08-NOV-2000; 2000US-0246477.
PR  08-NOV-2000; 2000US-0246478.
PR  08-NOV-2000; 2000US-0246523.
PR  08-NOV-2000; 2000US-0246524.
PR  08-NOV-2000; 2000US-0246525.
PR  08-NOV-2000; 2000US-0246526.
PR  08-NOV-2000; 2000US-0246527.
PR  08-NOV-2000; 2000US-0246528.
PR  08-NOV-2000; 2000US-0246532.
PR  08-NOV-2000; 2000US-0246609.
PR  08-NOV-2000; 2000US-0246610.
PR  08-NOV-2000; 2000US-0246611.
PR  08-NOV-2000; 2000US-0246613.
PR  17-NOV-2000; 2000US-0249207.
PR  17-NOV-2000; 2000US-0249208.
PR  17-NOV-2000; 2000US-0249209.
PR  17-NOV-2000; 2000US-0249210.
PR  17-NOV-2000; 2000US-0249211.
PR  17-NOV-2000; 2000US-0249212.
PR  17-NOV-2000; 2000US-0249213.
PR  17-NOV-2000; 2000US-0249214.
PR  17-NOV-2000; 2000US-0249215.
PR  17-NOV-2000; 2000US-0249216.
PR  17-NOV-2000; 2000US-0249217.
PR  17-NOV-2000; 2000US-0249218.
PR  17-NOV-2000; 2000US-0249244.
PR  17-NOV-2000; 2000US-0249245.
PR  17-NOV-2000; 2000US-0249264.
PR  17-NOV-2000; 2000US-0249265.
PR  17-NOV-2000; 2000US-0249297.
PR  17-NOV-2000; 2000US-0249299.
PR  17-NOV-2000; 2000US-0249300.
PR  01-DEC-2000; 2000US-0250160.
PR  01-DEC-2000; 2000US-0250391.
PR  05-DEC-2000; 2000US-0251030.
PR  05-DEC-2000; 2000US-0251988.
PR  05-DEC-2000; 2000US-0256719.
PR  06-DEC-2000; 2000US-0251479.
PR  08-DEC-2000; 2000US-0251856.
PR  08-DEC-2000; 2000US-0251868.
PR  08-DEC-2000; 2000US-0251869.
PR  08-DEC-2000; 2000US-0251989.
PR  08-DEC-2000; 2000US-0251990.
PR  11-DEC-2000; 2000US-0254097.
PR  05-JAN-2001; 2001US-0259678.
XX
PA  (HUMA-) HUMAN GENOME SCI INC.
XX
PI  Rosen CA,  Barash SC,  Ruben SM;
XX
DR  WPI; 2001-502630/55.
XX
PT  Polynucleotides encoding digestive system antigens, useful for
PT  diagnosing, treating, preventing and/or prognozing disorders of the
PT  digestive system, particularly cancer and cancer metastases -
XX
PS  Disclosure; SEQ ID NO 3342; 986pp; English.
XX
CC  The present invention provides the protein and coding sequences of a
CC  number of human digestive system antigens. These can be used in the
CC  diagnosis, treatment and prevention of digestive system disorders,
CC  including cancer, Meckel's diverticulum, bacterial or parasitic
CC  infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC  ulcerative colitis. The present sequence is a genomic DNA fragment
CC  encoding a digestive system antigen of the invention.
XX
SQ  Sequence 305 BP; 97 A; 52 C; 85 G; 71 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 22;  Length 305;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7578 ATCTCTTTAACCTGTTCTG 7596
             ||||||||||||||||||
Db        287 ATCTCTTTAACCTGTTCTG 305


RESULT 70
ABN20967
ID  ABN20967 standard; cDNA; 375 BP.
XX
AC  ABN20967;
XX
DT  24-JUN-2002  (first entry)
XX
```

```
DE   Human ORFX polynucleotide sequence SEQ ID NO:10411.
XX
KW   Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW   hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW   degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW   cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW   hypertension; hypothyroidism; cholesterol ester storage disease;
KW   immune deficiency; immune disorder; infectious disease;
KW   autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW   myasthenia gravis; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200192523-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US10836.
XX
PR   30-MAY-2000; 2000US-206132P.
PR   29-AUG-2000; 2000US-228716P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach MD;
XX
DR   WPI; 2002-106308/14.
DR   P-PSDB; ABP05215.
XX
PT   Novel human polypeptides and polynucleotides useful for diagnosing,
PT   preventing and treating cardiovascular disease, neurodegenerative,
PT   hyperproliferative disorders and autoimmune disorders -
XX
PS   Disclosure; SEQ ID 10411; 1037pp; English.
XX
CC   The present invention describes substantially purified human proteins
CC   (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC   in the specification). ABN15762 to ABN27252 encode the human ORFX
CC   proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC   treating or preventing a pathology associated with an ORFX-associated
CC   disorder in humans, and in the manufacture of a medicament for treating a
CC   syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC   sequences can be used in gene therapy. ORFX sequences can be used in the
CC   treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC   psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC   osteoarthritis, neurodegenerative disorders, disorders related to organ
CC   transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC   lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC   storage disease, various immune deficiencies and disorders, infectious
CC   diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC   arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC   disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC   useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC   bone degenerative disorders, or periodontal disease, and for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues and conditions resulting from
CC   systemic cytokine damage.
CC   N.B.  The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 375 BP; 169 A; 57 C; 58 G; 89 T; 2 other;

  Query Match             0.2%;  Score 19;  DB 24;  Length 375;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2383 AACAAGCAAATGGAAAACA 2401
             ||||||||||||||||||||
Db        296 AACAAGCAAATGGAAAACA 314

RESULT 71
ABX40398/c
ID   ABX40398 standard; cDNA; 387 BP.
XX
AC   ABX40398;
XX
DT   20-FEB-2003  (first entry)
XX
DE   Bovine EST associated with lactation/muscle/fat deposition #5563.
XX
KW   Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW   muscle deposition; fat deposition; genome mapping; gene identification;
KW   gene analysis; cattle breeding.
XX
OS   Bos Taurus.
XX
PN   US2002137139-A1.
XX
PD   26-SEP-2002.
XX
PF   24-SEP-2001; 2001US-0960352.
XX
PR   12-JAN-1999;   99US-115707P.
PR   11-JAN-2000; 2000US-0480902.
XX
PA   (BYAT/) BYATT J C.
PA   (MATH/) MATHIALAGAN N.
PA   (TAON/) TAO N.
PA   (WARR/) WARREN W C.
XX
PI   Byatt JC,  Mathialagan N,  Tao N,  Warren WC;
XX
DR   WPI; 2003-110599/10.
XX
PT   New nucleic acid associated with lactation, and muscle and fat
PT   deposition, useful for genome mapping, gene identification and
PT   analysis, cattle breeding, or for genetically improving cattle -
XX
PS   Claim 2; SEQ ID No 5563; 245pp; English.
XX
CC   The invention relates to a purified nucleic acid molecule associated with
CC   lactation or muscle and fat deposition (designated LMFD), derived
CC   from cattle, and the LMFD nucleic acid can specifically hybridise to a
CC   second nucleic acid molecule comprising any of 15112 nucleotide
CC   sequences, appearing as ABX34836-ABX49947, or complements of them.
CC   Also included are; (1) a transformed cell having a nucleic acid
CC   comprising an LMFD nucleic acid linked to a promoter and a 3' non-
CC   translated sequence that functions in the cell to cause termination of
CC   transcription and addition of polyadenylated ribonucleotides to a 3' end
CC   of the mRNA molecule; and (2) determining a level or pattern of a
CC   molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC   nucleic acid (comprising any of the 15112 nucleic acid sequences or its
CC   complement or fragment) with a complementary nucleic acid molecule
CC   obtained from the bovine cell or tissue, where hybridisation between the
CC   marker nucleic acid and the complementary nucleic acid permits the
CC   detection of the molecule; and (b) detecting the level or pattern of the
CC   complementary nucleic acid, where the detection of the complementary
CC   nucleic acid is predictive of the level or pattern of the molecule.
CC   The LMFD nucleic acid is used for determining a level or pattern
CC   of a molecule in a bovine cell or tissue. It is useful for genome
CC   mapping, gene identification and analysis, cattle breeding, preparation
CC   of constructs for use in cattle gene expression, or for genetically
CC   improving cattle. The present sequence is one of the 15112 bovine
CC   LMFD EST (expressed sequence tag) nucleic acids.
CC   Note: The present sequence was not shown in the specification but
CC   was obtained in electronic format from the USPTO web site:
CC   seqdata.uspto.gov/sequence.html?DocID=20020137139.
XX
SQ   Sequence 387 BP; 82 A; 90 C; 85 G; 130 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 25;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       4730 ATGTAATAAATAAACTTAA 4748
             ||||||||||||||||||
Db        335 ATGTAATAAATAAACTTAA 317


RESULT 72
ABX18291/c
ID   ABX18291 standard; cDNA; 401 BP.
XX
AC   ABX18291;
XX
DT   10-FEB-2003  (first entry)
XX
DE   Human GDP-mannose 4,6-dehydratase (GM4,6D) DNA #348.
XX
KW   Human; GDP-mannose 4,6-dehydratase; GM4,6D; gene; ss; inflammation;
KW   cellular fucosylation; glycoconjugate fucosylation; transplant rejection;
KW   arthritis; asthma; sepsis; reperfusion injury; stroke; infection;
KW   complex carbohydrate; gene replacement therapy; immunosuppressive;
KW   antiinflammatory; antiarthritic; antibacterial; cerebroprotective;
KW   antiasthmatic; vasotropic.
XX
OS   Homo sapiens.
XX
PN   US2002110548-A1.
XX
PD   15-AUG-2002.
XX
PF   11-JUN-2001; 2001US-0878574.
XX
PR   22-NOV-1996;   96US-0753233.
PR   03-DEC-1997;   97US-0984246.
PR   09-SEP-1998;   98US-0149674.
PR   14-JUN-1999;   99US-0333177.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Sullivan F,  Kriz R,  Kumar R;
XX
DR   WPI; 2003-066673/06.
XX
PT   New composition comprising GDP-mannose 4,6-dehydratase (GM4,6D)
PT   peptide, for manufacturing complex carbohydrates, or as targets for
PT   screening GM4,6D antagonists for treating e.g. arthritis, or transplant
PT   rejection  -
XX
PS   Disclosure; SEQ ID NO 350; 6pp; English.
XX
CC   The invention relates to a composition comprising a human GDP-mannose
CC   4,6-dehydratase (GM4,6D) peptide. The peptide is useful for identifying
CC   GM4,6D inhibitors. GM4,6D inhibitors are useful for reducing inflammation
CC   in a mammalian subject and for treating or ameliorating diseases affected
CC   by the level of cellular fucosylation or diseases affected by the
CC   fucosylation of glycoconjugates. These diseases include arthritis,
CC   transplant rejection, asthma, sepsis, reperfusion injury, stroke or
CC   infection. The GM4,6D peptide or a polynucleotide encoding it is also
CC   useful for manufacturing complex carbohydrates and as targets for
CC   screening small molecule antagonists of the activity of the enzyme. The
CC   polynucleotide is useful in developing an assay for defects in the
CC   enzyme, as well as in gene replacement therapy. Sequences
CC   ABX17942-ABX17944 and ABX17947-ABX33716 represent DNA molecules encoding
CC   human GM4,6D peptides of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 401 BP; 117 A; 65 C; 120 G; 99 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 25;  Length 401;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       102 CAGTCGCCTGTCAAATATA 120
             ||||||||||||||||||
Db        53 CAGTCGCCTGTCAAATATA 35


RESULT 73
AAH10172/c
ID   AAH10172 standard; cDNA; 564 BP.
XX
AC   AAH10172;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (3'-primer) SEQ ID NO:7007.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 3; SEQ ID 7007; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 564 BP; 153 A; 145 C; 130 G; 132 T; 4 other;

  Query Match             0.2%;  Score 19;  DB 22;  Length 564;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       7030 GGGCAGTACTGAGTTAATA 7048
             ||||||||||||||||||
Db         56 GGGCAGTACTGAGTTAATA 38


RESULT 74
ABQ26910/c
ID   ABQ26910 standard; DNA; 629 BP.
XX
AC   ABQ26910;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 13501.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 629 BP; 71 A; 75 C; 241 G; 242 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 24;  Length 629;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2369 ACTAAATAAATAAAAACAA 2387
              ||||||||||||||||||
Db        435 ACTAAATAAATAAAAACAA 417


RESULT 75
ABQ26911
ID   ABQ26911 standard; DNA; 629 BP.
XX
AC   ABQ26911;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 13502.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 629 BP; 242 A; 241 C; 75 G; 71 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 24;  Length 629;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2369 ACTAAATAAATAAAAACAA 2387
              ||||||||||||||||||
Db        195 ACTAAATAAATAAAAACAA 213


RESULT 76
```

```
AAI58779/c
ID   AAI58779 standard; cDNA; 719 BP.
XX
AC   AAI58779;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 982.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM39623.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 982; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 719 BP; 138 A; 249 C; 204 G; 128 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 22;  Length 719;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2883 CCTGACCAGGCGGGCGAGG 2901
             ||||||||||||||||||||
Db         62 CCTGACCAGGCGGGCGAGG 44
```

```
RESULT 77
AAH06747/c
ID   AAH06747 standard; cDNA; 816 BP.
XX
AC   AAH06747;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (5'-primer) SEQ ID NO:3582.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 1; SEQ ID 3582; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 816 BP; 243 A; 127 C; 162 G; 280 T; 4 other;

  Query Match             0.2%;  Score 19;  DB 22;  Length 816;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5144 AAATTAATTCCAGGGCTTG 5162
             ||||||||||||||||||||
```

```
Db        794 AAATTAATTCCAGGGCTTG 776


RESULT 78
AAC76215
ID   AAC76215 standard; cDNA; 818 BP.
XX
AC   AAC76215;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF1770 polynucleotide sequence SEQ ID NO:3539.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB42006.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 2703; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 818 BP; 170 A; 249 C; 237 G; 161 T; 1 other;

  Query Match             0.2%;  Score 19;  DB 21;  Length 818;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6263 TGACTGTCACCCCTACCCC 6281
             ||||||||||||||||||
Db        300 TGACTGTCACCCCTACCCC 318


RESULT 79
ABZ39135
ID   ABZ39135 standard; DNA; 888 BP.
XX
AC   ABZ39135;
XX
DT   07-MAR-2003  (first entry)
XX
DE   N. gonorrhoeae nucleotide sequence SEQ ID 2859.
XX
KW   Antibacterial; infection; vaccine; gene therapy; gene; ds.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO200279243-A2.
XX
PD   10-OCT-2002.
XX
PF   12-FEB-2002; 2002WO-IB02069.
XX
PR   12-FEB-2001; 2001GB-0003424.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Fontana MR,  Pizza M,  Masignani V,  Monaci E;
XX
DR   WPI; 2003-058415/05.
DR   P-PSDB; ABP78165.
XX
PT   New protein from Neisseria gonorrheae, useful for the manufacture of a
PT   medicament for treating or preventing N. gonorrheae infection  -
XX
PS   Disclosure; Page 397; 815pp; English.
XX
CC   The present invention relates to proteins from Neisseria gonorrhoeae.
CC   Also disclosed are the nucleic acid molecules encoding the proteins and
CC   antibodies that specifically bind to the proteins. The composition
CC   comprising the protein, nucleic acid or antibody is useful for the
CC   manufacture of a medicament for treating or preventing N. gonorrhoeae
CC   infection, this may be in the form of a vaccine or gene therapy.
CC   Sequences given in records ABZ37706-ABZ42016 represent nucleic acid
CC   molecules of the invention.
XX
SQ   Sequence 888 BP; 174 A; 233 C; 232 G; 249 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 25;  Length 888;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8304 GCCAATGTTTTATGGACAC 8322
             ||||||||||||||||||
Db        538 GCCAATGTTTTATGGACAC 556


RESULT 80
AAA65435
ID   AAA65435 standard; DNA; 900 BP.
XX
AC   AAA65435;
```

```
XX
DT   10-NOV-2000  (first entry)
XX
DE   Porcine BAC-PIGF2-1 contig 26.
XX
KW   Porcine; pig; wild boar; quantitative trait locus; QTL; chromosome 2;
KW   mapping; 2p1.7; select breeding; genotype; phenotype; muscle mass;
KW   fat deposition; IGF2; insulin-like growth factor 2; ds.
XX
OS   Sus scrofa.
XX
PN   WO200036143-A2.
XX
PD   22-JUN-2000.
XX
PF   16-DEC-1999;   99WO-EP10209.
XX
PR   16-DEC-1998;   98EP-0204291.
XX
PA   (UYLI-) UNIV LIEGE.
PA   (MELI-) MELICA HB.
PA   (SEGH-) SEGHERSGENTEC NV.
XX
PI   Andersson L,  Georges M,  Spincemaille G;
XX
DR   WPI; 2000-431612/37.
XX
PT   Selecting a domestic animal for having desired genotypic properties
PT   comprises testing the animal for the presence of a parentally imprinted
PT   quantitative trait locus which is related to muscle mass and/or fat
PT   deposition -
XX
PS   Example 3; Fig 6; 107pp; English.
XX
CC   The present invention describes a method (M1) for selecting a domestic
CC   animal for having desired genotypic properties. The method comprises
CC   testing the animal for the presence of a parentally imprinted
CC   quantitative trait locus (QTL). The pig QTL is located at chromosome 2,
CC   mapping at around position 2p1.7. Also described are: (1) an isolated
CC   and/or recombinant nucleic acid (N1) comprising a parentally imprinted
CC   QTL or its functional fragment; (2) an isolated and/or recombinant
CC   nucleic acid (N2) comprising a synthetic parentally imprinted QTL
CC   derived from at least one chromosome or its functional fragment; (3) an
CC   animal such as pig selected for having desired genotypic or potential
CC   phenotypic properties; (4) a transgenic animal comprising N1 or N2; and
CC   (5) sperm or an embryo derived from the animal of (3) or (4). N1 or its
CC   fragment is useful for selecting an animal destined for slaughter or a
CC   breeding animal having desired genotypic or potential phenotypic
CC   properties. The properties are related to muscle mass and/or fat
CC   deposition. The sperm or an embryo are useful in breeding animals
CC   destined for slaughter. AAA65418 to AAA65524 represent contigs 1 to 10
CC   and 19 to 115 which were isolated from porcine BAC-PIGF2-1 which
CC   contains the INS and IGF2 (insulin-like growth factor) genes. These
CC   sequences where used in an example from the present invention for
CC   generating a reference sequence of IGF2 and flanking loci in the pig.
XX
SQ   Sequence 900 BP; 220 A; 236 C; 228 G; 216 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 21;  Length 900;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2956 CATCTATAAAAAAATCGTC 2974
              ||||||||||||||||||||
Db        815 CATCTATAAAAAAATCGTC 833


RESULT 81
ABX76413/c
ID   ABX76413 standard; DNA; 948 BP.
XX
AC   ABX76413;
XX
DT   02-APR-2003  (first entry)
XX
DE   Lung cancer-associated polynucleotide #277.
XX
KW   Lung cancer-associated polynucleotide; gene; ds; cytostatic; emphysema;
KW   antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
KW   small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
KW   chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
KW   interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
XX
OS   Unidentified.
XX
PN   WO200286443-A2.
XX
PD   31-OCT-2002.
XX
PF   18-APR-2002; 2002WO-US12476.
XX
PR   18-APR-2001; 2001US-284770P.
PR   10-MAY-2001; 2001US-290492P.
PR   09-NOV-2001; 2001US-339245P.
PR   13-NOV-2001; 2001US-350666P.
PR   29-NOV-2001; 2001US-334370P.
PR   12-APR-2002; 2002US-372246P.
XX
PA   (EOSB-) EOS BIOTECHNOLOGY INC.
XX
PI   Aziz N,  Murray R;
XX
DR   WPI; 2003-093161/08.
DR   P-PSDB; ABU56684.
XX
PT   Detecting a lung cancer-associated transcript in a cell from a patient
PT   for treating lung cancer, by contacting a biological sample from the
PT   patient with a polynucleotide that exhibits increased or decreased
PT   expression in lung cancer  -
XX
PS   Claim 22; Page 400-401; 453pp; English.
XX
CC   The invention relates to a method for detecting a lung cancer-associated
CC   transcript in a cell from a patient, comprising contacting a biological
CC   sample from the patient with a polynucleotide that selectively hybridises
CC   to a sequence that is at least 80 % identical to a gene that exhibits
CC   increased or decreased expression in lung cancer samples. Lung
CC   cancer-associated polynucleotides and polypeptides are used for
CC   identifying a compound that modulates a lung cancer-associated
CC   polypeptide, for inhibiting proliferation of a lung cancer-associated
CC   cell to treat lung cancer in a patient and for treating a mammal having
CC   lung cancer by administering a modulatory compound identified. The
CC   methods are useful for treating lung cancer, such as small cell lung
CC   cancer, non-small cell lung cancer or other benign or precancerous
CC   lesions, e.g. atelectasis, emphysema, bronchitis, chronic obstructive
CC   pulmonary disease, fibrosis, hypersensitivity pneumonitis, interstitial
CC   pulmonary fibrosis, asthma and bronchiectasis. The genes, polynucleotides
CC   and polypeptides are useful for diagnostic purposes and as targets for
CC   screening for therapeutic compounds that modulate lung cancer, such as
CC   antibodies. Sequences ABX76124-ABX76474 represent lung cancer-associated
CC   polynucleotides of the invention.
XX
SQ   Sequence 948 BP; 227 A; 240 C; 276 G; 205 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 25;  Length 948;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3426 GGAGTGGCCGGCGGGGTCC 3444
              ||||||||||||||||||||
Db        522 GGAGTGGCCGGCGGGGTCC 504


RESULT 82
AAT30244
ID   AAT30244 standard; DNA; 989 BP.
```

```
XX
AC   AAT30244;
XX
DT   25-MAR-2003  (updated)
DT   16-DEC-1996  (first entry)
XX
DE   Cotton fibre cell-specific cDNA clone CKFB10-A11.
XX
KW   Cotton fibre; promoter; differential screening; leaf; ovule; root;
KW   flower; PCR; polymerase chain reaction; homology; transgenic plant; ds.
XX
OS   Gossypium hirsutum.
XX
PN   US5521078-A.
XX
PD   28-MAY-1996.
XX
PF   19-OCT-1994;   94US-0298687.
XX
PR   18-MAY-1992;   92US-0885970.
PR   04-OCT-1988;   88US-0253243.
PR   21-NOV-1990;   90US-0617239.
PR   19-OCT-1994;   94US-0298687.
XX
PA   (CETU ) AGRACETUS INC.
XX
PI   John M;
XX
DR   WPI; 1996-267794/27.
XX
PT   Isolation of fibre-specific cotton promoter sequences - using
PT   selected DNA probes to screen genomic DNA fragments, for production
PT   of cotton fibres with improved characteristics
XX
PS   Claim 1; Column 31-34; 46pp; English.
XX
CC   Cotton fibre cell-specific promoter sequences were isolated by
CC   differential screening of a cotton plant cDNA library.  Of 4788 clones
CC   from a 10 day cell library screened with leaf cDNAs, 800 clones not
CC   present in the leaf were isolated. These were screened with cDNAs from
CC   ovule, root and flower mRNAs and resulted in 79 clones isolated.  PCR
CC   analysis was then used to remove cross-hybridising clones.  This
CC   resulted in the isolation of 18 cDNA clones specifically expressed in
CC   cotton fibre cells (AAT30242-4 and AAT30253-67). These cDNAs were then
CC   used to screen for homologous genomic sequences (AAT30245-53 and
CC   AAT30268) in order to obtain the corresp. promoter sequences.
CC   This cDNA clone contains an insert of 1 kb which hybridised with RNAs of
CC   1.1 and 0.9 kb.  This clone is fibre-specific.
CC   The promoters isolated from the fibre cell-specific clones can be used
CC   to generate transgenic cotton plants and lines producing fibres having
CC   altered quantity and quality.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 989 BP; 362 A; 197 C; 214 G; 216 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 17;  Length 989;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1359 AAATGGAAAAACAAAAAAA 1377
              |||||||||||||||||||
Db        400 AAATGGAAAAACAAAAAAA 418


RESULT 83
AAT13038
ID   AAT13038 standard; cDNA; 989 BP.
XX
AC   AAT13038;
XX
DT   25-MAR-2003  (updated)
DT   27-MAY-1996  (first entry)
XX
DE   Cotton fibre-specific cDNA clone A11.
XX
KW   Cotton; fibre; promoter; transgenic plant; crop improvement; ds.
XX
OS   Gossypium hirsutum strain Coker 312.
XX
PN   US5495070-A.
XX
PD   27-FEB-1996.
XX
PF   18-MAY-1992;   92US-0885970.
XX
PR   18-MAY-1992;   92US-0885970.
PR   04-OCT-1988;   88US-0253243.
PR   21-NOV-1990;   90US-0617239.
XX
PA   (CETU ) AGRACETUS INC.
XX
PI   John M;
XX
DR   WPI; 1996-139095/14.
XX
PT   New isolated fibre-specific promoters - used for introducing
PT   altered fibre-specific characteristics into plants, partic. cotton.
XX
PS   Example 3; Column 31-34; 48pp; English.
XX
CC   Cotton fibre-specific cDNA clone A11 (AAT13038) was isolated from a
CC   cDNA library of cotton var. Coker 312 10-day-old boll cells by a
CC   subtractive hybridization procedure.  It hybridizes to RNAs of 1.1
CC   and 0.9 kb.  A11 and other fibre-specific cDNA clones (see AAT13033-37
CC   and AAT13039-T13050) were used to screen cotton genomic libraries,
CC   leading to the isolation of genomic clones (see AAT13025-32 and
CC   AAT13052-53) contg. sequences capable of promoting gene expression in
CC   fibre cells.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 989 BP; 362 A; 197 C; 214 G; 216 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 17;  Length 989;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1359 AAATGGAAAAACAAAAAAA 1377
              |||||||||||||||||||
Db        400 AAATGGAAAAACAAAAAAA 418


RESULT 84
AAT70045
ID   AAT70045 standard; cDNA; 989 BP.
XX
AC   AAT70045;
XX
DT   25-MAR-2003  (updated)
DT   20-AUG-1997  (first entry)
XX
DE   Cotton fibre specific cDNA clone A11.
XX
KW   cotton; E6; fibre; promoter; transgenic plant; truncated;
KW   heterologous gene expression; ds.
XX
OS   Gossypium hirsutum strain Coker 312.
XX
PN   US5620882-A.
XX
PD   15-APR-1997.
XX
PF   19-OCT-1994;   94US-0298829.
XX
PR   18-MAY-1992;   92US-0885970.
PR   04-OCT-1988;   88US-0253243.
PR   21-NOV-1990;   90US-0617239.
```

```
PR   19-OCT-1994;   94US-0298829.
XX
PA   (CETU ) AGRACETUS INC.
XX
PI   John M;
XX
DR   WPI; 1997-235185/21.
XX
PT   DNA constructs contg. truncated promoter sequence  - for
PT   fibre-specific gene expression in cotton plants
XX
PS   Example 3; Column 31-34; 48pp; English.
XX
CC   AAT70040-57 are cotton fibre-specific cDNA clones which can be used to
CC   obtain genomic clones containing fibre-specific promoters. Claimed DNA
CC   constructs comprise a truncated promoter sequence (from one of
CC   AAT70031-38) that promotes preferential gene expression in plant fibre
CC   cells, a protein coding sequence not naturally associated with the
CC   promoter sequence and a 3' termination sequence. The DNA constructs are
CC   useful for expressing foreign genes in fibre-producing plants, esp. to
CC   produce transgenic cotton plants with varied cotton fibre characteristics
CC   and quality. The present sequence comprises A11 cDNA isolated from clone
CC   CKFB10-A11 (CK = Coker; FB10 = 10 day old bolls).
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 989 BP; 362 A; 197 C; 214 G; 216 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 18;  Length 989;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1359 AAATGGAAAAACAAAAAAA 1377
              |||||||||||||||||||
Db        400 AAATGGAAAAACAAAAAAA 418


RESULT 85
AAT62614
ID   AAT62614 standard; cDNA to mRNA; 989 BP.
XX
AC   AAT62614;
XX
DT   25-MAR-2003  (updated)
DT   14-MAY-1997  (first entry)
XX
DE   Cotton fibre specific cDNA clone CKFB10-A11.
XX
KW   cotton; fibre-specific; strength; transgenic plant; anthesis;
KW   developmentally regulated; E6; H6; antisense; sense; ss.
XX
OS   Gossypium hirsutum strain Coker 312.
XX
PN   US5597718-A.
XX
PD   28-JAN-1997.
XX
PF   20-SEP-1995;   95US-0530797.
XX
PR   20-SEP-1995;   95US-0530797.
PR   04-OCT-1988;   88US-0253243.
PR   21-NOV-1990;   90US-0617239.
PR   18-OCT-1993;   93US-0138814.
XX
PA   (CETU ) AGRACETUS.
XX
PI   Brill WJ,  John ME,  Umbeck PF;
XX
DR   WPI; 1997-108326/10.
XX
PT   Prodn. of transgenic cotton plants - by transformation with the H6
PT   coding sequence or E6 anti-sense sequence, produces fibre of altered
PT   strength
XX
PS   Example 4; Column 35-36; 33pp; English.
XX
CC   AAT62609-24 are cotton fibre-specific cDNA clones which can be used to
CC   identify genomic clones. This clone, CKFB10-A11, is fibre-specific. (CK
CC   = Coker; FB = Fibre; 10, 15 or 23 = age in days of fibre cells; A1 and
CC   the last character and number stand for clone identity). The
CC   fibre-specific genes were identified by differential cDNA library
CC   screenings. Coding sequences from these isolated genes are used in sense
CC   or antisense orientation to alter the fibre characteristics, e.g.
CC   strength, of transgenic fibre-producing plants.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 989 BP; 362 A; 197 C; 214 G; 216 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 18;  Length 989;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1359 AAATGGAAAAACAAAAAAA 1377
              |||||||||||||||||||
Db        400 AAATGGAAAAACAAAAAAA 418


RESULT 86
AAZ35548
ID   AAZ35548 standard; DNA; 989 BP.
XX
AC   AAZ35548;
XX
DT   28-JAN-2000  (first entry)
XX
DE   cDNA sequence a cotton fibre gene from clone CKB10-A11.
XX
KW   Clone CKB10-A11; cotton plant; fibre gene; transgenic plant; ss;
KW   cellulose synthesis; fibre development.
XX
OS   Gossypium hirsutum.
XX
PN   US5981834-A.
XX
PD   09-NOV-1999.
XX
PF   27-JAN-1997;   97US-0787335.
XX
PR   21-NOV-1990;   90US-0617239.
PR   20-SEP-1995;   95US-0530797.
PR   04-OCT-1988;   88US-0253243.
PR   18-OCT-1993;   93US-0138814.
XX
PA   (MONS ) MONSANTO CO.
XX
PI   Umbeck PF,  Brill WJ,  John ME;
XX
DR   WPI; 2000-012274/01.
XX
PT   New transgenic cotton plants containing foreign fibre-specific DNA to
PT   alter fibre characteristics  -
XX
PS   Claim 1; Column 35-36; 32pp; English.
XX
CC   This is the nucleotide sequence of clone CKB10-A11, which represents a
CC   cotton fibre gene. This nucleotide sequence is used in the invention
CC   which relates to the production of new transgenic cotton plants
CC   containing DNA that is fibre specific (i.e. sequences AAZ35544-Z35548)
CC   for producing fibre with altered characteristics. The foreign DNA
CC   express proteins contained in the cotton plants which influence fibre
CC   development, fibre characteristics and cellulose synthesis in cotton to
CC   provide plants and lines with varied cotton fibre characteristics and
CC   quality.
XX
SQ   Sequence 989 BP; 362 A; 197 C; 214 G; 216 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 21;  Length 989;
```

```
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1359 AAATGGAAAAACAAAAAAA 1377
              |||||||||||||||||||
Db        400 AAATGGAAAAACAAAAAAA 418


RESULT 87
AAA59190/c
ID   AAA59190 standard; DNA; 1056 BP.
XX
AC   AAA59190;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Endogenetic retroviral fragment integrated into the human genome.
XX
KW   Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
KW   gag gene; pregnancy; multiple sclerosis; T cell proliferation;
KW   multiple sclerosis associated retrovirus; MSRV; ss.
XX
OS   Homo sapiens.
XX
PN   WO200043521-A2.
XX
PD   27-JUL-2000.
XX
PF   21-JAN-2000; 2000WO-FR00144.
XX
PR   21-JAN-1999;   99FR-0000888.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Paranhos-Baccala G,  Mallet F,  Voisset C;
XX
DR   WPI; 2000-499229/44.
XX
PT   New nucleic acid from human endogenous retrovirus, useful e.g. for
PT   diagnosis of autoimmune disease and complications of pregnancy,
PT   contains at least part of the gag gene  -
XX
PS   Claim 4; Page 36-37; 53pp; French.
XX
CC   The present sequence represents an endogenetic retroviral nucleic acid
CC   fragment, which is associated with an autoimmune disease, and is
CC   integrated into the human genome. The fragment is originally derived
CC   from a multiple sclerosis associated retrovirus (MSRV), which is
CC   homologous to a novel retrovirus, human endogenous retrovirus W (HERV-W).
CC   The fragment comprises (or consists of) at least part of the gag gene of
CC   MSRV. HERV-W retrovirus is associated with autoimmune disease,
CC   failure of pregnancy or disorders of pregnancy. The nucleic acid
CC   fragment, or proteins derived from it, are useful for diagnosis of
CC   autoimmune disease (specifically multiple sclerosis) and for monitoring
CC   pregnancy. The nucleic acid fragments may also be used for in situ
CC   labelling of isolated chromosomes, while the transcription product can be
CC   used to study or monitor T cell proliferation in vitro.
XX
SQ   Sequence 1056 BP; 344 A; 247 C; 221 G; 244 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 21;  Length 1056;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        377 ACATCAATATAGCCGTCAG 395
              |||||||||||||||||||
Db        557 ACATCAATATAGCCGTCAG 539


RESULT 88
AAX29726/c
ID   AAX29726 standard; DNA; 1186 BP.
XX
AC   AAX29726;
XX
DT   08-JUN-1999  (first entry)
XX
DE   Insert from pET21C-clone 2 from MSRV-1.
XX
KW   Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
KW   rheumatoid polyarthritis; ss.
XX
OS   Multiple sclerosis related virus type 1.
XX
PN   FR2765588-A1.
XX
PD   08-JAN-1999.
XX
PF   07-JUL-1997;   97FR-0008816.
XX
PR   07-JUL-1997;   97FR-0008816.
XX
PA   (INMR ) BIO MERIEUX.
XX
DR   WPI; 1999-098275/09.
DR   P-PSDB; AAW99558.
XX
PT   Nucleic acid sequences of retrovirus called MSRV-1 - associated with
PT   multiple sclerosis or rheumatoid polyarthritis
XX
PS   Disclosure; Fig 8; 83pp; French.
XX
CC   This sequence represents the insert found in the plasmid pET21C-clone 2
CC   containing a fragment from a novel multiple sclerosis related virus
CC   type 1 (MSRV1).  The sequence can be used in diagnostic, prophylactic or
CC   therapeutic compositions to inhibit expression of a multiple sclerosis
CC   related virus and/or virus associated with rheumatoid polyarthritis.
XX
SQ   Sequence 1186 BP; 378 A; 284 C; 260 G; 264 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 20;  Length 1186;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        377 ACATCAATATAGCCGTCAG 395
              |||||||||||||||||||
Db        596 ACATCAATATAGCCGTCAG 578


RESULT 89
AAX29725/c
ID   AAX29725 standard; DNA; 1247 BP.
XX
AC   AAX29725;
XX
DT   08-JUN-1999  (first entry)
XX
DE   Insert from pET28C-clone 2 from MSRV-1.
XX
KW   Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
KW   rheumatoid polyarthritis; ss.
XX
OS   Multiple sclerosis related virus type 1.
XX
PN   FR2765588-A1.
XX
PD   08-JAN-1999.
XX
PF   07-JUL-1997;   97FR-0008816.
XX
PR   07-JUL-1997;   97FR-0008816.
XX
PA   (INMR ) BIO MERIEUX.
XX
DR   WPI; 1999-098275/09.
DR   P-PSDB; AAW99557.
```

```
XX
PT   Nucleic acid sequences of retrovirus called MSRV-1 - associated with
PT   multiple sclerosis or rheumatoid polyarthritis
XX
PS   Disclosure; Fig 7; 83pp; French.
XX
CC   This sequence represents the insert found in the plasmid pET28C-clone 2
CC   containing a fragment from a novel multiple sclerosis related virus
CC   type 1 (MSRV1).  The sequence can be used in diagnostic, prophylactic or
CC   therapeutic compositions to inhibit expression of a multiple sclerosis
CC   related virus and/or virus associated with rheumatoid polyarthritis.
XX
SQ   Sequence 1247 BP; 391 A; 305 C; 276 G; 273 T; 2 other;

  Query Match             0.2%;  Score 19;  DB 20;  Length 1247;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        377 ACATCAATATAGCCGTCAG 395
              ||||||||||||||||||| 
Db        656 ACATCAATATAGCCGTCAG 638


RESULT 90
ABX71022
ID   ABX71022 standard; cDNA; 1287 BP.
XX
AC   ABX71022;
XX
DT   05-MAR-2003  (first entry)
XX
DE   Novel human cDNA sequence #247.
XX
KW   Human; gene; ss; nervous system disorder; peripheral neuropathy;
KW   Huntington's disease; amyotrophic lateral sclerosis; haemophilia;
KW   neurodegenerative disease; Parkinson's disease; Alzheimer's disease;
KW   autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;
KW   insulin-dependent diabetes mellitus; anaemia; thrombocytopaenia; wound;
KW   ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;
KW   fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;
KW   coagulation disorder; cancer; tumour; inflammatory disease;
KW   septic shock; Crohn's disease; anaphylaxis; proliferation; chemotactic;
KW   differentiation; stem cell growth factor; haematopoiesis; chemokinetic;
KW   haemostatic; antiinflammatory; expressed sequence tag; EST.
XX
OS   Homo sapiens.
XX
PN   WO200281731-A2.
XX
PD   17-OCT-2002.
XX
PF   29-JAN-2002; 2002WO-US01222.
XX
PR   30-JAN-2001; 2001US-0774528.
XX
PA   (HYSE-) HYSEQ INC.
PA   (GOOD/) GOODRICH R W.
XX
PI   Tang TY,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Wehrman T,  Wang J,  Wang D,  Drmanac RT;
XX
DR   WPI; 2003-058563/05.
XX
PT   Novel polypeptide useful for treating neurodegenerative diseases,
PT   myeloid or lymphoid cell disorders, bone disorders, mechanical and
PT   traumatic disorders, coagulation disorders, and inflammatory diseases
PT   -
XX
PS   Claim 1; Page -; 612pp; English.
XX
CC   This invention relates to the cDNA sequences encoding an isolated
CC   novel human polypeptide. The protein encoded by the nucleic acid of
CC   the invention is useful for treating central and peripheral nervous
CC   system diseases (e.g. peripheral neuropathy, Huntington's disease,
CC   amyotrophic lateral sclerosis); neurodegenerative diseases (e.g.
CC   Parkinson's disease, Alzheimer's disease); autoimmune disease (e.g.
CC   systemic lupus erythematosus, rheumatoid arthritis, insulin-dependent
CC   diabetes mellitus); myeloid or lymphoid cell disorders (e.g. anaemia
CC   and thrombocytopaenia); wounds, ulcers, burns; bone disorders (e.g.
CC   osteoporosis, osteoarthritis); mechanical and traumatic disorders (e.g.
CC   stroke, head trauma); lung or liver fibrosis; reperfusion injury in
CC   various tissues; bacterial, viral or fungal infections; allergic
CC   conditions such as allergic rhinitis, asthma; coagulation disorders
CC   (e.g. haemophilia); cancer and tumours; and inflammatory diseases (e.g.
CC   septic shock, Crohn's disease, anaphylaxis). The protein may be used to
CC   inhibit the growth, infection or function of infectious agents such as
CC   bacteria, fungi, viruses, or to effect bodily characteristics,
CC   biorhythms or circadian cycles of rhythms. The protein may also
CC   have proliferation/differentiation, stem cell growth factor,
CC   haematopoiesis regulation, immune stimulating or suppressing,
CC   chemotactic/chemokinetic, haemostatic and thrombolytic, receptor/ligand,
CC   and antiinflammatory activities. The cDNA sequences of the invention are
CC   useful for expressing recombinant protein for analysis. The present
CC   sequence represents a novel human cDNA sequence of the invention,
CC   this sequence is an expressed sequence tag (EST) and was identified
CC   using subtractive hybridisation.
XX
SQ   Sequence 1287 BP; 355 A; 309 C; 334 G; 289 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 25;  Length 1287;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1357 TTAAATGGAAAAACAAAAA 1375
              ||||||||||||||||||| 
Db       1232 TTAAATGGAAAAACAAAAA 1250


RESULT 91
AAC50413
ID   AAC50413 standard; DNA; 1295 BP.
XX
AC   AAC50413;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 64748.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
```

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

```
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             0.2%;  Score 19;  DB 21;  Length 1295;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        130 CCACCGGAAATGGGTTGGT 148
              ||||||||||||||||||| 
Db        148 CCACCGGAAATGGGTTGGT 166


RESULT 92
AAC37673
ID   AAC37673 standard; DNA; 1298 BP.
XX
AC   AAC37673;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 18250.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
```

PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.2%; Score 19; DB 21; Length 1298;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 CCACCGGAAATGGGTTGGT 148
||||||||||||||||||||
Db 150 CCACCGGAAATGGGTTGGT 168

RESULT 93
AAX29772/c
ID AAX29772 standard; DNA; 1400 BP.
XX
AC AAX29772;
XX
DT 22-JUN-1999 (first entry)
XX
DE L.lactis HsdS gene #1.
XX
KW HsdR; HsdM; HsdS; restriction/modification; bacteriophage; resistance;
KW lactic acid bacterium; ds.
XX
OS Lactococcus lactis.
XX
PN FR2767831-A1.
XX
PD 05-MAR-1999.
XX
PF 02-SEP-1997; 97FR-0010885.
XX
PR 02-SEP-1997; 97FR-0010885.
XX
PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX
PI Chopin MC, Clier F, Ehrlich SD, Gautier M, Schouler C;
XX
DR WPI; 1999-183265/16.
DR P-PSDB; AAY04662.
XX
PT Bacteriophage resistance mechanism subunit polypeptides - of lactic
PT acid bacteria, especially Lactococcus lactis
XX
PS Claim 5; Page 44-46; 65pp; French.
XX
CC The invention relates to polypeptides that constitute the HsdR, HsdM or
CC HsdS subunit of a R/M [restriction/modification] type Ic bacteriophage
CC resistance mechanism active against the phages of lactic acid bacteria,
CC especially Lactococcus lactis. This sequence represents the gene
CC encoding the Lactococcus lactis HsdS subunit. Nucleic acids encoding
CC the HsdR, HsdM or HsdS fragment can be used for expression of at least
CC one bacteriophage resistance mechanism in a lactic acid bacterium.
XX
SQ Sequence 1400 BP; 502 A; 196 C; 279 G; 423 T; 0 other;

Query Match 0.2%; Score 19; DB 20; Length 1400;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4604 CTAAATTTGTTAAATGAAA 4622
||||||||||||||||||||
Db 1248 CTAAATTTGTTAAATGAAA 1230

```
RESULT 94
AAC50427
ID   AAC50427 standard; DNA; 1408 BP.
XX
AC   AAC50427;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 64803.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
```

```
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             0.2%;  Score 19;  DB 21;  Length 1408;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        130 CCACCGGAAATGGGTTGGT 148
              ||||||||||||||||||
Db        171 CCACCGGAAATGGGTTGGT 189


RESULT 95
AAQ32365/c
ID   AAQ32365 standard; DNA; 1412 BP.
XX
AC   AAQ32365;
XX
DT   25-MAR-2003  (updated)
DT   22-APR-1993  (first entry)
XX
DE   MAGE-9 genomic DNA.
XX
KW   Melanoma antigen; MAGE TRA; melanoma antigen tumor rejection antigen;
KW   tumor rejection antigen precursor; MAGE; antigen E; gene family; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             428..1375
FT                   /*tag=  a
XX
PN   WO9220356-A1.
XX
PD   26-NOV-1992.
XX
PF   22-MAY-1992;   92WO-US04354.
XX
PR   23-MAY-1991;   91US-0705702.
PR   09-JUL-1991;   91US-0728838.
PR   23-SEP-1991;   91US-0764364.
PR   12-DEC-1991;   91US-0807043.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon T,  Chomez P,  De Plaen E,  Lurquin C,  Traversari C;
PI   Van Den Eynde B,  Van Der Bruggen P,  Van Pel A;
XX
DR   WPI; 1992-415460/50.
XX
PT   Nucleic acid mol. encoding a human tumour rejection antigen
PT   precursor - useful as an immunostimulant in a vaccine for
PT   treating and preventing cancers, also useful in diagnosis
XX
PS   Disclosure; Page 91; 142pp; English.
XX
CC   The sequences given in AAQ32352-69 represent a new family of genes
CC   refered to as melanoma antigens (MAGE).  The cDNAs of this gene
CC   family were identified during the isolation of the antigen E gene.
CC   The MAGE cDNAs, when tested, did not transfer expression of antigen
CC   E, but they did show substantial homology to the antigen E cDNA
CC   sequence.  The MAGE DNAs share a certain degree of homology with each
CC   other and are expressed in tumour cells including several types of
CC   human tumor cells as well as in human tumors.  MAGE expression is not
CC   restricted to melanomas.  MAGE refers to a family of tumor rejection
CC   antigen precursors.  The antigens resulting from these genes are
CC   referred to as MAGE TRAs or melanoma antigen tumor rejection antigens.
CC   See also AAQ32351.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 1412 BP; 324 A; 400 C; 395 G; 293 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 13;  Length 1412;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3426 GGAGTGGCCGGCGGGGTCC 3444
              ||||||||||||||||||
Db        949 GGAGTGGCCGGCGGGGTCC 931


RESULT 96
AAQ72490/c
ID   AAQ72490 standard; DNA; 1412 BP.
XX
AC   AAQ72490;
XX
DT   25-MAR-2003  (updated)
DT   22-JUN-1995  (first entry)
XX
```

```
DE   Tumour rejection antigen MAGE-9 genomic DNA.
XX
KW   Tumour rejection antigen; melanoma antigen-9; MAGE-9; MAGE-3;
KW   cancer; cytolytic T cells; antigen D; human leucocyte antigen;
KW   ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             428..1375
FT                   /*tag= a
XX
PN   WO9423031-A1.
XX
PD   13-OCT-1994.
XX
PF   17-MAR-1994;   94WO-US02877.
XX
PR   26-MAR-1993;   93US-0037230.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon-falleur T,  Gaugler B,  Van DEN EYNDE B,  Van DER BRUGGEN P;
XX
DR   WPI; 1994-333192/41.
XX
PT   New tumour rejection antigen precursor MAGE3 - useful in
PT   treatment and diagnosis of cancer
XX
PS   Example 33; Page 76; 105pp; English.
XX
CC   AAQ72490 is the genomic DNA which contains the coding sequence for
CC   melanoma antigen-9 (MAGE-9). Another melanoma antigen MAGE-3 is
CC   encoded by AAQ72470, this is a tumour rejection antigen precursor.
CC   Melanomas characterised by the expression of MAGE-3 can be detected,
CC   or monitored, by contacting a test sample with an agent that can
CC   recognise MAGE-3. The melanoma can be treated by the administration
CC   of cytolytic T cells specific for the complex of antigen D (the
CC   mature rejection antigen derived from MAGE-3) and a human leucocyte
CC   antigen (esp. HLA-A1).
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 1412 BP; 324 A; 402 C; 393 G; 293 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 15;  Length 1412;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3426 GGAGTGGCCGGCGGGGTCC 3444
              ||||||||||||||||||
Db        949 GGAGTGGCCGGCGGGGTCC 931


RESULT 97
AAT01169/c
ID   AAT01169 standard; DNA; 1412 BP.
XX
AC   AAT01169;
XX
DT   25-MAR-2003  (updated)
DT   26-FEB-1996  (first entry)
XX
DE   MAGE-9 gene.
XX
KW   MAGE-9; melanoma; tumour rejection antigen; cancer; diagnosis; ss.
XX
OS   Homo sapiens.
XX
PN   WO9523874-A1.
XX
PD   08-SEP-1995.
XX
PF   23-FEB-1995;   95WO-US02203.
XX
PR   30-NOV-1994;   94US-0346774.
PR   01-MAR-1994;   94US-0204727.
PR   10-MAR-1994;   94US-0209172.
PR   01-SEP-1994;   94US-0299849.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   De Plaen E,  Boon-falleur T,  Lethe B,  Szikora J,  De Smet C;
PI   Chomez P,  Gaugler B,  Van Den Eynde B,  Brasseur F,  Patard J;
PI   Weynants P,  Marchand M,  Van Der Bruggen P;
XX
DR   WPI; 1995-320586/41.
XX
PT   Determn. of cancerous condition(s) - using a nucleic acid as a
PT   primer to determine expression of a MAGE tumour rejection antigen
PT   precursor
XX
PS   Example 33; Page 82-83; 121pp; English.
XX
CC   A family of human tumour rejection antigen precursor, MAGE, genes
CC   (AAT05091-99, AAT01165-71) was isolated from various tumour cell lines.
CC   The MAGE 7, 8, 9, 10 and 11 genes (AAT01167-71) were obtd. from
CC   cosmid libraries from PHA-activated lymphocytes of melanoma
CC   patient MZ2.  MAGE serve as markers for tumour diagnosis.  The
CC   genes are silent in normal cells.
CC   (Updated on 25-MAR-2003 to correct PI field.)
XX
SQ   Sequence 1412 BP; 324 A; 402 C; 393 G; 293 T; 0 other;

  Query Match             0.2%;  Score 19;  DB 16;  Length 1412;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3426 GGAGTGGCCGGCGGGGTCC 3444
              ||||||||||||||||||
Db        949 GGAGTGGCCGGCGGGGTCC 931


RESULT 98
AAX84125/c
ID   AAX84125 standard; DNA; 1412 BP.
XX
AC   AAX84125;
XX
DT   08-SEP-1999  (first entry)
XX
DE   MAGE-9 gene.
XX
KW   Tumour rejection antigen; vaccine; cancer; MAGE-9 gene; ss.
XX
OS   Homo sapiens.
XX
PN   US5925729-A.
XX
PD   20-JUL-1999.
XX
PF   02-MAY-1994;   94US-0142368.
XX
PR   02-MAY-1994;   94US-0142368.
PR   23-MAY-1991;   91US-0705702.
PR   09-JUL-1991;   91US-0728838.
PR   23-SEP-1991;   91US-0764365.
PR   12-DEC-1991;   91US-0807043.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon T,  Chomez P,  De Plaen E,  Lurquin C,  Traversari C;
PI   Van Den Eynde B, Van Der Bruggen P,  Van Pel A;
XX
DR   WPI; 1999-418294/35.
XX
PT   New tumour rejection antigen is useful as a vaccine against
```

PT cancerous diseases
XX
PS Disclosure; Column 75-76; 58pp; English.
XX
CC This sequence represents the MAGE-9 gene sequence.
CC The invention relates to a tumour rejection antigen sequence that is
CC useful as a tumour rejection antigen for vaccination against cancerous
CC conditions.
XX
SQ Sequence 1412 BP; 324 A; 402 C; 393 G; 293 T; 0 other;

Query Match 0.2%; Score 19; DB 20; Length 1412;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3426 GGAGTGGCCGGCGGGGTCC 3444
||||||||||||||||||||
Db 949 GGAGTGGCCGGCGGGGTCC 931

RESULT 99
AAC40335
ID AAC40335 standard; DNA; 1412 BP.
XX
AC AAC40335;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 27931.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.

```
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match              0.2%;  Score 19;  DB 21;  Length 1412;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        130 CCACCGGAAATGGGTTGGT 148
              ||||||||||||||||||| 
Db        173 CCACCGGAAATGGGTTGGT 191


RESULT 100
AAA64206
ID   AAA64206 standard; DNA; 1456 BP.
XX
AC   AAA64206;
XX
DT   20-DEC-2000  (first entry)
XX
DE   Complete sequence of the muscle creatine kinase gene promoter region.
XX
KW   Zebrafish; promoter; cytokeratin gene; muscle creatine kinase gene;
KW   fast skeletal muscle isoform; myosin light chain 2 gene; fish embryo;
KW   acidic ribosomal protein gene; structural gene; skin epithelia;
KW   chimeric gene; transgenic fish; steroid hormone; heavy metal;
KW   environmental monitoring; ornamental fish; ss.
XX
OS   Danio rerio.
XX
FH   Key             Location/Qualifiers
FT   TATA_signal      1389..1394
FT                    /*tag= a
XX
PN   WO200049150-A1.
XX
PD   24-AUG-2000.
XX
PF   16-JUL-1999;   99WO-SG00079.
XX
PR   18-FEB-1999;   99SG-0000811.
PR   14-JUL-1999;   99SG-9999999.
XX
PA   (UYSI-) UNIV SINGAPORE NAT.
PA   (XUYY/) XU Y.
PA   (HEJJ/) HE J.
PA   (YANT/) YAN T.
XX
PI   Gong Z,  Lam TJ,  Ju B;
XX
DR   WPI; 2000-558297/51.
XX
PT   Zebrafish gene promoters, used to create transgenic zebrafish, useful
PT   for sensing steroid hormones and heavy metals in water samples -
XX
PS   Claim 14; Page 36-37; 72pp; English.
XX
CC   AAA64205-07 and AAA64217 represent the promoters of Zebrafish
CC   cytokeratin gene, muscle creatine kinase gene, fast skeletal muscle
CC   isoform of the myosin light chain 2 gene, and acidic ribosomal protein
CC   gene, repectively. The promoters of these genes are muscle- and
CC   skin-specific. They are capable of directing a structural gene to be
CC   predominantly expressed in skin epithelia when it is inserted in front of
CC   the structural gene and introduced into fish embryos. The promoters can
CC   be used in recombinant DNA molecules, or chimeric genes and the chimeric
CC   genes can be used to create transgenic fish or in a vector used to
CC   transform a cell. The transgenic fish can be used in methods for
CC   detecting steroid hormones, or heavy metals such as zinc, copper,
CC   cadmium and mercury, in a water sample (i.e. in environmental monitoring
CC   and for monitoring pollution). The transgenic fish may also be used as
CC   ornamental fish.
XX
```

SQ Sequence 1456 BP; 438 A; 298 C; 300 G; 420 T; 0 other;

Query Match 0.2%; Score 19; DB 21; Length 1456;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1402 TACTTTGAGTTCAAAAGAA 1420
              ||||||||||||||||||
Db        638 TACTTTGAGTTCAAAAGAA 656
```

Search completed: August 22, 2003, 01:24:40
Job time : 2082 secs